EXPERT JURNAL MANAJEMEN SISTEM INFORMASI DAN TEKNOLOGI

Algoritma K-Means Clustering Untuk Segmentasi Produk Kosmetik Terlaris Berdasarkan Volume Penjualan

Nasya Salsabila [1], **Tukino**[2], **Agustia Hananto**[3], **Bayu Priyatna**[4]
[1234]Sistem Informasi, Ilmu Komputer, Universitas Buana Perjuangan, Karawang, Indonesia
[1] si.22nasyasalsabila@mhs.ubpkarawang.ac.id, [2] tukino@ubpkarawang.ac.id, [3] agustia.hananto@ubpkarawang.ac.id, [4] bayu.priyatna@ubpkarawang.ac.id

ABSTRACT – *Wardah, as one of the prominent local cosmetic brands, has succeeded in attracting consumer interest by presenting quality products, halal certified, and affordable prices. With the increasing competition in this industry, it is important for companies to utilize sales data more effectively to design more precise business and marketing strategies. Analyzing sales data based on products sold is essential to understand market dynamics and consumer preferences. This study utilizes the K-Means Clustering algorithm to segment products based on sales data, focusing on two main variables, namely the number of units sold and the total sales value. The results of determining the optimal number of clusters using the Elbow and Davies-Bouldin Index (DBI) methods show that five clusters provide the best data separation and compactness, with the lowest DBI value of 0.7772. Further analysis of the clusters shows that Cluster 0 is the largest group, covering 65 products or around 46.76% of the total. Products in this cluster have a low average price, which is around IDR 38,733, with total sales reaching more than IDR 31 million. Meanwhile, cluster 1 consists of only 8 products (5.76%), but has the highest average price, which is IDR 227,000. Although the number of sales is smaller, this cluster indicates the presence of premium or exclusive products. Cluster 2 includes 40 products (28.78%) with an average price of around IDR 100,016 and total sales reaching almost IDR 49 million. Cluster 3 stands out with a high average number of variants (6 variants per product), the highest total sales of around IDR 68.5 million, and the largest sales volume with 1,073 units sold. This cluster consists of 23 products (16.55%) and shows that although the number of products is smaller, its contribution to total sales is very significant. Finally, cluster 4 is the smallest cluster with only 3 products (2.16%), but has a very high number of variants, reaching an average of 20.67 variants per product, and a mid-to-high price of around IDR 88,781. Overall, the results of this clustering provide important insights for business strategies, such as pricing, stock management, and promotional targeting based on the characteristics of each product group.*

Keywords: ***Davies-Bouldin Index; K-Means Clustering; Metode Elbow; Wardah.***

ABSTRAK – Wardah, sebagai salah satu merek kosmetik lokal yang menonjol, berhasil menarik minat konsumen dengan menghadirkan produk-produk berkualitas, bersertifikat halal, serta memiliki harga yang terjangkau. Dengan semakin ketatnya persaingan dalam industri ini, penting bagi perusahaan untuk memanfaatkan data penjualan secara lebih efektif guna merancang strategi bisnis dan pemasaran yang lebih presisi. Menganalisis data penjualan berdasarkan produk yang terjual sangat penting untuk memahami dinamika pasar dan preferensi konsumen. Penelitian ini memanfaatkan algoritma K-Means Clustering untuk melakukan segmentasi produk berdasarkan data penjualan, dengan fokus pada dua variabel utama, yaitu jumlah unit yang terjual dan nilai total penjualan. Hasil penentuan jumlah klaster optimal menggunakan metode Elbow dan Davies-Bouldin Index (DBI) menunjukkan bahwa lima klaster memberikan pemisahan dan kekompakan data terbaik, dengan nilai DBI terendah sebesar 0,7772. Analisis lebih lanjut terhadap klaster menunjukkan bahwa Klaster 0 merupakan kelompok terbesar, mencakup 65 produk atau sekitar 46,76% dari total. Produk dalam klaster ini memiliki harga rata-rata yang rendah, yaitu sekitar Rp38.733, dengan total penjualan mencapai lebih dari Rp31 juta. Sementara itu, klaster 1 terdiri dari hanya 8 produk (5,76%), namun memiliki harga rata-rata paling tinggi, yakni Rp227.000. Meski jumlah penjualannya lebih kecil, klaster ini mengindikasikan keberadaan produk-produk premium atau eksklusif. Klaster 2 mencakup 40 produk (28,78%) dengan harga rata-rata sekitar Rp100.016 dan total penjualan mencapai hampir Rp49 juta. Klaster 3 menonjol dengan jumlah varian rata-rata yang tinggi (6 varian per produk), total penjualan tertinggi yaitu sekitar Rp68,5 juta, dan volume penjualan terbesar dengan 1.073 unit terjual. Klaster ini terdiri dari 23 produk (16,55%) dan menunjukkan bahwa meskipun jumlah produknya lebih sedikit, kontribusinya terhadap total penjualan sangat signifikan. Terakhir, klaster 4 merupakan klaster terkecil dengan hanya 3 produk (2,16%), namun memiliki jumlah varian sangat tinggi, rata-rata mencapai 20,67 varian per produk, dan harga menengah ke atas sekitar Rp88.781. Secara keseluruhan, hasil klasterisasi ini memberikan wawasan penting untuk strategi bisnis, seperti penentuan harga, pengelolaan stok, dan penargetan promosi berdasarkan karakteristik masing-masing kelompok produk.

Kata Kunci: Davies-Bouldin Index; K-Means Clustering; Metode Elbow; Wardah.

This work is licensed under a Creative Commons Attribution 4.0 International License

DOI http://dx.doi.org/10.36448/expert.v15i1.4318
e-ISSN 2745-7265 p-ISSN 2088-5555 EXPERT Vol. 15 No. 1
June 11, 2025 – Hal.42

1. PENDAHULUAN

Industri kosmetik di Indonesia telah mengalami pertumbuhan yang signifikan dalam beberapa tahun terakhir, didorong oleh meningkatnya kesadaran masyarakat terhadap pentingnya perawatan diri dan penampilan. Produk kosmetik kini tidak lagi dianggap sebagai kebutuhan pelengkap, melainkan telah menjadi bagian dari gaya hidup sehari-hari, terutama di kalangan wanita muda dan professional [1]. Wardah, sebagai salah satu merek kosmetik lokal yang menonjol, berhasil menarik minat konsumen dengan menghadirkan produk-produk berkualitas, bersertifikat halal, serta memiliki harga yang terjangkau [2]. Dengan semakin ketatnya persaingan dalam industri ini, penting bagi perusahaan untuk memanfaatkan data penjualan secara lebih efektif guna merancang strategi bisnis dan pemasaran yang lebih presisi.

Menganalisis data penjualan berdasarkan produk yang terjual sangat penting untuk memahami dinamika pasar dan preferensi konsumen [3]. Dengan mengelompokkan produk berdasarkan karakteristik penjualan, perusahaan dapat mengidentifikasi pola-pola tersembunyi yang tidak terlihat melalui analisis konvensional. Pengelompokan produk berdasarkan perilaku pelanggan dapat mengungkap struktur kategorisasi produk yang lebih mendalam, memberikan wawasan tambahan bagi pengambilan keputusan bisnis [4].

Salah satu pendekatan yang dapat digunakan untuk menganalisis data penjualan adalah dengan teknik clustering, yang bertujuan untuk mengelompokkan objek-objek berdasarkan kemiripan karakteristik [5]. Dalam konteks ini, pengelompokan produk berdasarkan volume penjualan dapat memberikan wawasan mendalam mengenai performa produk dan preferensi konsumen [6]. Metode clustering memungkinkan perusahaan untuk mengidentifikasi kelompok produk yang berkinerja tinggi, sedang, dan rendah, sehingga strategi pemasaran dan distribusi dapat disesuaikan [7].

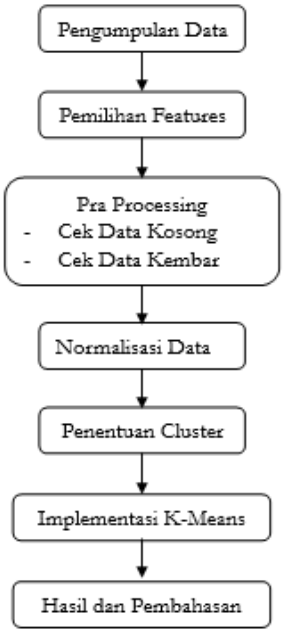
Metode K-means merupakan algoritma clustering yang paling umum digunakan karena kesederhanaannya, kecepatan dalam komputasi, serta kemampuannya dalam mengelompokkan data ke dalam kluster yang homogen [8]. Algoritma ini bekerja dengan menentukan sejumlah pusat kluster (centroid) secara acak, lalu mengelompokkan data berdasarkan kedekatan jarak ke centroid tersebut [9]. Proses ini diulangi hingga pusat kluster tidak lagi berubah secara signifikan. Dalam implementasi pada data penjualan, fitur-fitur seperti volume penjualan bulanan, rata-rata penjualan, dan frekuensi pembelian dapat digunakan sebagai variabel input. K-means terbukti efektif dalam mengelompokkan calon potensial berdasarkan data transaksi yang serupa [10].

Dalam penelitian ini, data yang digunakan terdiri dari sepuluh produk terlaris Wardah berdasarkan volume penjualan dalam periode tertentu. Analisis dilakukan dengan menerapkan algoritma K-means untuk mengelompokkan produk, metode elbow untuk menentukan jumlah kluster optimal, dan Davies-Bouldin Index sebagai alat evaluasi kualitas klaster. Dengan pendekatan ini, diharapkan dapat dihasilkan segmentasi produk yang akurat yang dapat menjadi dasar dalam strategi pemasaran, penataan stok, dan pengembangan produk.

Dengan memahami pola distribusi dan kinerja penjualan produk melalui teknik clustering, perusahaan seperti Wardah dapat mengambil keputusan berbasis data yang lebih tepat sasaran. Pendekatan ini tidak hanya meningkatkan efisiensi operasional, tetapi juga memungkinkan perusahaan untuk merespons perubahan pasar dengan lebih adaptif. Hasil dari studi ini dapat menjadi acuan bagi pelaku industri kosmetik lainnya dalam mengintegrasikan teknologi analitik ke dalam proses bisnis mereka.

2. METODOLOGI

Penelitian ini menggunakan metode deskriptif kuantitatif, yaitu pendekatan yang bertujuan untuk mendeskripsikan pola atau karakteristik data secara numerik melalui analisis statistic [11]. Metode ini memungkinkan peneliti untuk menggambarkan fenomena berdasarkan data yang terukur tanpa melakukan manipulasi variabel. Selanjutnya, proses pengumpulan data dilakukan dengan menghimpun dataset yang relevan sesuai dengan tujuan penelitian. Data ini dapat diperoleh dari sumber primer yang merupakan catatan transaksi penjualan kosmetik Wardah dalam bulan Maret 2025.



Gambar 1. Alur Penelitian

This work is licensed under a Creative Commons Attribution 4.0 International License
DOI http://dx.doi.org/10.36448/expert.v15i1.4318
e-ISSN 2745-7265 p-ISSN 2088-5555 EXPERT Vol. 15 No. 1
June 11, 2025 – Hal.43

Setelah data terkumpul, dilakukan pemilihan fitur (*features selection*) untuk menentukan atribut-atribut yang relevan dan berpengaruh terhadap hasil analisis. Proses ini bertujuan untuk menyaring variabel penting guna meningkatkan akurasi dan efisiensi dalam tahap selanjutnya [12]. Kemudian, dilakukan pra-pemrosesan data yang mencakup pemeriksaan data kosong (missing values) dan data duplikat. Langkah ini penting untuk memastikan kualitas dan konsistensi data sebelum digunakan dalam pemodelan. Data yang tidak lengkap atau berulang dapat mengganggu proses analisis dan menghasilkan hasil yang tidak valid [13].

Untuk menentukan jumlah cluster yang optimal dalam analisis klaster, digunakan metode Elbow untuk menentukan jumlah cluster yang paling sesuai dalam algoritma seperti K-Means. Pendekatan ini dilakukan dengan menghitung nilai *Within-Cluster Sum of Squares* (WCSS) pada berbagai jumlah cluster, kemudian memetakan hasilnya dalam bentuk grafik. Titik optimal terlihat pada bagian siku dari kurva, yaitu saat penurunan nilai WCSS mulai melambat meskipun jumlah cluster terus bertambah. Titik tersebut menunjukkan jumlah cluster yang paling efisien dalam merepresentasikan pola dalam data [14].

Kualitas cluster yang terbentuk dievaluasi menggunakan *Davies-Bouldin Index* (DBI), yang merupakan metrik evaluasi internal yang digunakan untuk menilai kualitas hasil klasterisasi dalam algoritma seperti K-Means [15]. DBI mengukur rata-rata kemiripan antara setiap klaster dengan klaster yang paling mirip dengannya, berdasarkan rasio antara jarak intra-klaster (dalam klaster) dan jarak antar-klaster (antar pusat klaster). Semakin rendah nilai DBI, semakin baik kualitas klasterisasi, karena menunjukkan bahwa klaster lebih kompak dan terpisah dengan baik satu sama lain. DBI memberikan nilai dari 0 hingga tak hingga, dengan nilai yang lebih rendah menunjukkan klaster yang lebih baik. Nilai DBI yang rendah menunjukkan bahwa klaster memiliki sebaran internal yang kecil dan terpisah dengan baik dari klaster lain. Sebaliknya, nilai DBI yang tinggi menunjukkan bahwa klaster memiliki sebaran internal yang besar dan/atau tidak terpisah dengan baik dari klaster lain [16].

Tahap akhir adalah penerapan algoritma K-means, yaitu metode pembelajaran tanpa pengawasan yang mengelompokkan data ke dalam beberapa cluster berdasarkan kedekatan nilai fitur. Proses ini melibatkan pemilihan centroid awal, penghitungan jarak tiap data ke centroid, dan pembaruan posisi centroid hingga konvergen [17]. Hasil dari K-means digunakan untuk mengidentifikasi pola-pola dalam data yang dapat dijadikan dasar dalam pengambilan keputusan atau analisis lanjutan.

3. HASIL DAN PEMBAHASAN

Hasil

Data yang digunakan dalam penelitian ini mencakup beberapa atribut penting, yaitu kode produk, nama produk, varian atau jenis warna (*shade*), isi atau volume produk, harga satuan, jumlah produk yang terjual, serta nilai total penjualan dengan total transaksi sebanyak 394 transaksi. Setiap baris data mencerminkan transaksi tertentu, di mana satu jenis produk dapat tercatat lebih dari sekali karena perbedaan dalam ukuran atau jumlah pembelian. Untuk tujuan segmentasi menggunakan algoritma K-Means Clustering, dilakukan pengelompokan awal berdasarkan nama produk (nm_produk). Hal ini didasarkan bahwa nm_produk mencerminkan entitas unik dari produk yang dipasarkan. Dengan melakukan agregasi berdasarkan nama produk, peneliti dapat menurunkan sejumlah fitur numerik penting seperti jumlah varian, harga rata-rata produk, jumlah penjualan produk, dan total penjualan produk. Fitur-fitur inilah yang selanjutnya menjadi variabel masukan dalam proses klasterisasi. Pendekatan ini bertujuan menghasilkan segmentasi produk yang merepresentasikan performa penjualan secara keseluruhan, bukan berdasarkan transaksi tunggal, sehingga lebih tepat digunakan dalam penyusunan strategi pemasaran dan pengambilan keputusan bisnis.

Tabel 1. Hasil Grouping Data Berdasarkan Nama Produk

No	Nama Produk	Varian	Harga	Jumlah	Total
0	10% Vitamin C Ferulic Acid Radiant Charge Serum	1	120000	7	840000
1	14X Hyaluron + Pentavitin Deep Water Restore G...	1	119000	4	476000
2	Acne Face Powder	1	54000	10	540000
3	Acne Treatment Gel	1	30000	8	240000
4	Acnederm Acne Care Serum	1	75000	14	1050000
...	...	...	...	...	...
134	Vitamin C + Adenosine Radiant Charge Gel Moist...	1	119000	5	595000
135	Wardah Eyebrow Pencil	1	44000	13	572000
136	Wardah Eyeliner Pencil	1	44500	15	667500
137	Wardah Perfect Bright Moisturizer Bright + Oil...	1	30800	8	246400
138	Wardah UV Shield Airy Smooth Sunscreen Serum S...	2	48250	43	2085500

This work is licensed under a Creative Commons Attribution 4.0 International License
DOI http://dx.doi.org/10.36448/expert.v15i1.4318
e-ISSN 2745-7265 p-ISSN 2088-5555 EXPERT Vol. 15 No. 1
June 11, 2025 – Hal.44

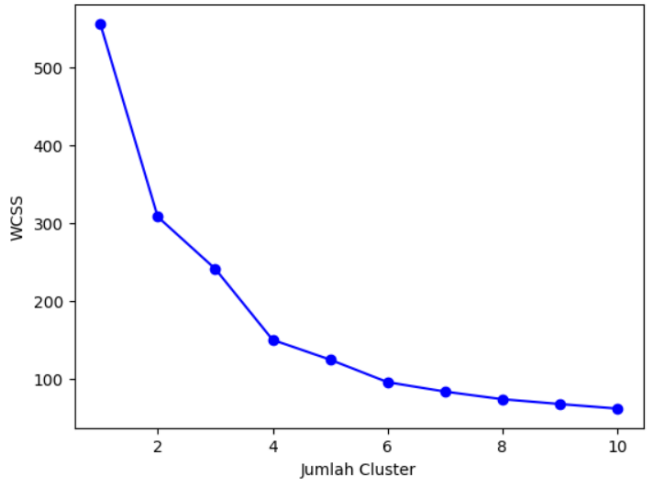
Tabel tersebut merupakan hasil grouping berdasarkan nama produk. Kolom "Nama Produk" berisi nama lengkap dari setiap produk. Kolom Varian menunjukkan jumlah variasi atau edisi dari produk tersebut, yang sebagian besar bernilai 1, namun ada juga yang lebih dari satu, menandakan kemungkinan perbedaan ukuran atau formula. Harga satuan dari setiap produk tercantum di kolom "Harga", sedangkan kolom "Jumlah" menunjukkan banyaknya unit produk yang terjual. Total penjualan dihitung dari hasil perkalian antara harga dan jumlah unit, dan nilainya tercatat pada kolom "Total". Sebagai contoh, produk 10% Vitamin C Ferulic Acid Radiant Charge Serum dijual seharga Rp120.000 per unit dan terjual sebanyak 7 unit, menghasilkan total penjualan sebesar Rp840.000.

Hasil dari pemilihan features akan dilakukan pengecekan untuk data kosong dan data kembar, berdasarkan hasil pengujian, tidak ditemukan data kosong dan data kembar. Langkah selanjutnya adalah melakukan normalisasi data menggunakan metode StandardScaler.

Tabel 2. Hasil Normalisasi Data Penjualan Wardah

No	Varian	Harga	Jumlah	Total
0	-0.406594	0.926828	-0.699048	-0.330973
1	-0.406594	0.907248	-0.869021	-0.586391
2	-0.406594	-0.365438	-0.529076	-0.541483
3	-0.406594	-0.835353	-0.642391	-0.751992
4	-0.406594	0.045737	-0.302445	-0.183617
...	...	...	...	...
134	-0.406594	0.907248	-0.812363	-0.502889
135	-0.406594	-0.561236	-0.359103	-0.519028
136	-0.406594	-0.551446	-0.245788	-0.452016
137	-0.406594	-0.819689	-0.642391	-0.747501
138	-0.132242	-0.478022	1.340624	0.542992

Setelah data dinormalisasi akan ditentukan untuk jumlah kluster optimal menggunakan metode Ellbow.



Gambar 2. Grafik Penentuan Kluster Menggunakan Metode Elbow

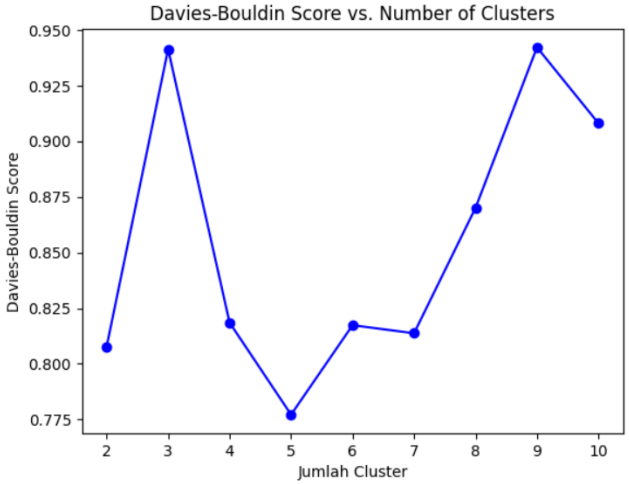
Grafik yang ditampilkan pada gambar 1 merupakan Elbow Method yang digunakan untuk menentukan jumlah klaster optimal dalam algoritma K-Means Clustering. Sumbu horizontal menunjukkan jumlah klaster (dari 1 hingga 10), sedangkan sumbu vertikal menunjukkan nilai WCSS (*Within-Cluster Sum of Squares*), yaitu total jarak kuadrat antara titik data dan pusat klasternya. Tujuannya adalah untuk mencari titik di mana penurunan WCSS mulai melambat secara signifikan, yang menunjukkan bahwa penambahan klaster setelah titik tersebut tidak memberikan perbaikan yang berarti dalam pemodelan.

Dalam grafik ini, penurunan tajam pada WCSS terjadi dari klaster 1 hingga sekitar klaster 4. Setelah klaster ke-4, penurunan WCSS menjadi lebih landai. Titik "siku" atau "elbow" tersebut sekitar klaster ke-4 menunjukkan jumlah klaster yang ideal, karena setelah titik itu, penambahan klaster tidak secara signifikan menurunkan WCSS. Dengan kata lain, membagi data

This work is licensed under a
Creative Commons Attribution 4.0 International License
DOI http://dx.doi.org/10.36448/expert.v15i1.4318
e-ISSN 2745-7265 p-ISSN 2088-5555 EXPERT Vol. 15 No. 1
June 11, 2025 – Hal.44

menjadi empat klaster memberikan keseimbangan yang baik antara kompleksitas model dan kualitas klasterisasi.

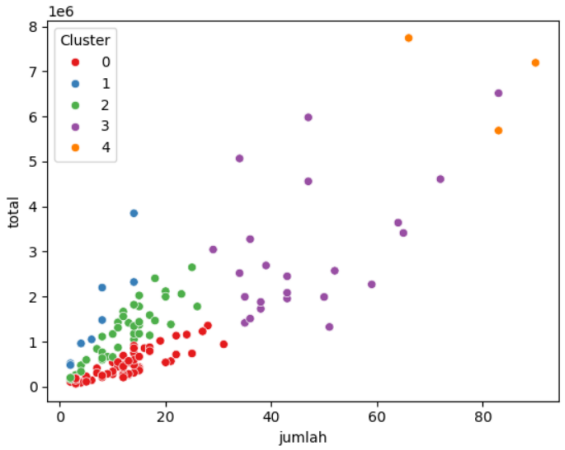
Hasil penentuan kluster dengan metode Elbow akan dievaluasi lagi menggunakan metode *Davies-Bouldin Index* (DBI).



Gambar 3. Grafik Evaluasi Kluster Menggunakan *Davies Bouldin Score*

Grafik *Davies-Bouldin Index* (DBI) yang ditampilkan menggambarkan hubungan antara jumlah klaster dan nilai Davies-Bouldin Score. Metode ini digunakan untuk mengevaluasi kualitas klaster berdasarkan seberapa baik klaster dipisahkan dan seberapa kompak masing-masing klaster. Nilai DBI yang lebih rendah menunjukkan hasil klasterisasi yang lebih optimal. Dari grafik tersebut terlihat bahwa nilai terendah terjadi pada saat jumlah klaster adalah lima dengan nilai evaluasi sebesar 0,7772, yang mengindikasikan bahwa pemisahan dan kepadatan antar klaster berada pada kondisi terbaiknya ketika data dibagi menjadi lima kelompok.

Jika dibandingkan dengan grafik Elbow sebelumnya, yang menggunakan metode Within-Cluster Sum of Squares (WCSS), titik "siku" atau peralihan penurunan signifikan terjadi pada jumlah klaster empat dengan nilai evaluasi sebesar 0,8185. Berdasarkan hasil tersebut pada penelitian ini akan menggunakan jumlah kluster lima berdasrakan hasil evaluasi DBI. Langkah selanjutnya adalah menerapkan kkluster optimal dalam K-means.



Gambar 4. Hasil Klustering Menggunakan K-Means dengan 5 Kluster

Hasil penelitian ini menunjukkan penerapan metode K-Means Clustering dalam melakukan segmentasi produk berdasarkan data penjualan, dengan mempertimbangkan dua variabel utama yaitu jumlah unit terjual (jumlah) dan total pendapatan penjualan (total). Hasil visualisasi menunjukkan adanya distribusi klaster yang cukup jelas. Klaster 0 (ditandai warna merah) berisi produk-produk dengan jumlah penjualan dan total pendapatan yang rendah, mencerminkan produk dengan performa terendah. Klaster 1 (biru) terdiri atas produk dengan jumlah penjualan yang tidak terlalu tinggi, tetapi sebagian memiliki total penjualan yang cukup besar, kemungkinan karena harga jual per unit yang tinggi. Klaster 2 (hijau) menunjukkan kelompok produk dengan kinerja sedang baik dari sisi jumlah maupun total penjualan. Sementara itu, Klaster 3 (ungu) mencakup produk dengan jumlah dan total penjualan tinggi, yang dapat dikategorikan sebagai produk unggulan atau best-seller. Adapun Klaster 4 (oranye) mencerminkan produk dengan performa penjualan yang sangat tinggi baik dari

This work is licensed under a
Creative Commons Attribution 4.0 International License
_DOI http://dx.doi.org/10.36448/expert.v15i1.4318
e-ISSN 2745-7265 p-ISSN 2088-5555 EXPERT Vol. 15 No. 1
June 11, 2025 – Hal.2

sisi kuantitas maupun nilai, meskipun jumlah produknya relatif sedikit, sehingga klaster ini kemungkinan mewakili produk premium atau sangat populer.

Tabel 3. Hasil Tiap Kluster

Cluster	Mean Varian	Mean Harga	Sum Jumlah	Sum Total	Count Produk	Persentase
0	1.292307692	38733.58974	822	31652600	65	46.76%
1	1	227000	58	12864000	8	5.76%
2	1.325	100015.8333	496	48685200	40	28.78%
3	6	65201.88406	1073	68504500	23	16.55%
4	20.66666667	88780.95238	239	20616500	3	2.16%

Tabel yang ditampilkan merupakan hasil segmentasi produk menggunakan metode klasterisasi, di mana produk dikelompokkan menjadi lima klaster berdasarkan rata-rata varian, harga, jumlah unit terjual, total penjualan, dan jumlah produk di setiap kelompok. Klaster 0 merupakan kelompok terbesar, mencakup 65 produk atau sekitar 46,76% dari total. Produk dalam klaster ini memiliki harga rata-rata yang rendah, yaitu sekitar Rp38.733, dengan total penjualan mencapai lebih dari Rp31 juta. Ini menunjukkan bahwa produk-produk dalam klaster ini cenderung terjangkau dan memiliki volume penjualan tinggi, menjadikannya produk massal atau populer di pasaran.

Sementara itu, klaster 1 terdiri dari hanya 8 produk (5,76%), namun memiliki harga rata-rata paling tinggi, yakni Rp227.000. Meski jumlah penjualannya lebih kecil, klaster ini mengindikasikan keberadaan produk-produk premium atau eksklusif. Klaster 2 mencakup 40 produk (28,78%) dengan harga rata-rata sekitar Rp100.016 dan total penjualan mencapai hampir Rp49 juta. Produk-produk dalam klaster ini dapat dikategorikan sebagai kelas menengah dari sisi harga dan performa penjualan. Klaster 3 menonjol dengan jumlah varian rata-rata yang tinggi (6 varian per produk), total penjualan tertinggi yaitu sekitar Rp68,5 juta, dan volume penjualan terbesar dengan 1.073 unit terjual. Klaster ini terdiri dari 23 produk (16,55%) dan menunjukkan bahwa meskipun jumlah produknya lebih sedikit, kontribusinya terhadap total penjualan sangat signifikan—kemungkinan terdiri dari produk-produk dengan banyak pilihan dan permintaan tinggi.

Terakhir, klaster 4 merupakan klaster terkecil dengan hanya 3 produk (2,16%), namun memiliki jumlah varian sangat tinggi, rata-rata mencapai 20,67 varian per produk, dan harga menengah ke atas sekitar Rp88.781. Produk-produk dalam klaster ini kemungkinan besar bersifat sangat spesifik atau menyasar pasar niche, dengan banyak pilihan varian namun volume penjualan yang lebih terbatas. Secara keseluruhan, hasil klasterisasi ini memberikan wawasan penting untuk strategi bisnis, seperti penentuan harga, pengelolaan stok, dan penargetan promosi berdasarkan karakteristik masing-masing kelompok produk.

Berdasarkan hasil klasterisasi dan analisis data penjualan, pengambilan 10 produk terbaik didasarkan pada total unit terjual tertinggi. Pemilihan ini bertujuan untuk mengidentifikasi produk-produk yang paling diminati oleh konsumen dan memberikan kontribusi besar terhadap volume penjualan secara keseluruhan. Produk-produk dengan jumlah unit terjual terbanyak umumnya berasal dari klaster dengan karakteristik harga yang terjangkau dan varian yang tidak terlalu banyak, seperti klaster 0 dan 3. Hal ini menunjukkan bahwa produk yang mudah dijangkau oleh pasar dan memiliki keberagaman terbatas cenderung lebih cepat terserap oleh konsumen. Dengan fokus pada 10 produk terlaris, perusahaan dapat merumuskan strategi yang lebih tepat sasaran dalam hal promosi, distribusi, dan pengelolaan stok. Selain itu, analisis ini juga dapat digunakan sebagai acuan untuk pengembangan produk serupa yang berpotensi sukses di pasar.

Tabel 4. 10 Produk Teratas Berdasar Jumlah Penjualan

Nama Produk	Varian	Harga	Jumlah	Total	Cluster
Colorfit Velvet Matte Lip Mousse	22	79900	90	7191000	4
Colorfit Velvet Powder Foundation	8	78500	83	6515500	3
Matte Lip Cream	26	68500	83	5685500	4
Colorfit Fresh Matte Lip Ink	8	64000	72	4608000	3
Colorfit Perfect Glow Cushion	14	117942.9	66	7740000	4
Colorfit Velvet Powder Foundation Refill	7	52500	65	3412500	3
Lightening Two Way Cake Light Feel	7	56900	64	3641600	3
Refill Lightening Two Way Cake Light Feel	7	38500	59	2271500	3

This work is licensed under a Creative Commons Attribution 4.0 International License
DOI http://dx.doi.org/10.36448/expert.v15i1.4318
e-ISSN 2745-7265 p-ISSN 2088-5555 EXPERT Vol. 15 No. 1
June 11, 2025 – Hal.47

Colorfit Ultralight Matte Lipstick	5	49500	52	2574000	3
C-Defense Vitamin C + Panthenol Anti-Pollution Radiant Creamy Wash	1	26000	51	1326000	3

Berdasarkan hasil analisis terhadap sepuluh produk terbaik berdasarkan jumlah penjualan, dapat disimpulkan bahwa produk-produk yang memiliki tingkat penjualan tertinggi didominasi oleh kategori produk bibir dan wajah dari lini Colorfit. Produk dengan jumlah penjualan tertinggi adalah Colorfit Velvet Matte Lip Mousse sebanyak 90 unit, diikuti oleh Colorfit Velvet Powder Foundation dan Matte Lip Cream, masing-masing sebanyak 83 unit. Secara umum, produk-produk dalam daftar ini memiliki harga jual yang bervariasi, mulai dari produk dengan harga terjangkau seperti C-Defense Vitamin C + Panthenol Creamy Wash seharga Rp26.000 hingga produk premium seperti Colorfit Perfect Glow Cushion dengan harga rata-rata lebih dari Rp117.000.

Dari sisi segmentasi klaster berdasarkan algoritma K-Means, sebagian besar produk dalam daftar ini termasuk dalam Cluster 3, yang mencerminkan produk dengan kinerja penjualan tinggi namun belum masuk kategori tertinggi dalam hal total pendapatan. Sementara itu, Cluster 4 diisi oleh produk-produk dengan performa penjualan yang sangat unggul, baik dari jumlah maupun nilai total penjualan, seperti Colorfit Velvet Matte Lip Mousse, Matte Lip Cream, dan Colorfit Perfect Glow Cushion. Hal ini menunjukkan bahwa Cluster 4 merepresentasikan produk-produk unggulan yang layak dijadikan prioritas dalam strategi pemasaran dan pengelolaan persediaan, sedangkan Cluster 3 mencakup produk dengan potensi kuat untuk dikembangkan lebih lanjut.

Pembahasan

Penelitian ini memanfaatkan algoritma K-Means Clustering untuk melakukan segmentasi produk berdasarkan data penjualan, dengan fokus pada dua variabel utama, yaitu jumlah unit yang terjual dan nilai total penjualan. Data awal dikelompokkan berdasarkan nama produk untuk memastikan bahwa proses klasterisasi merepresentasikan performa produk secara keseluruhan. Hasil penentuan jumlah klaster optimal menggunakan metode Elbow dan Davies-Bouldin Index (DBI) menunjukkan bahwa lima klaster memberikan pemisahan dan kekompakan data terbaik, dengan nilai DBI terendah sebesar 0,7772 [18].

Analisis lebih lanjut terhadap klaster menunjukkan bahwa Klaster 0 mencakup sebagian besar produk dengan harga terjangkau dan volume penjualan tinggi, mengindikasikan karakteristik produk massal. Sebaliknya, Klaster 1 berisi produk-produk berharga tinggi dengan volume penjualan lebih rendah, kemungkinan merupakan produk premium. Klaster 3, yang mencakup produk-produk dengan banyak varian dan total penjualan tertinggi, serta Klaster 4, yang berisi produk dengan performa penjualan terbaik secara keseluruhan, menjadi sorotan utama dalam strategi pemasaran.

Pemilihan sepuluh produk terlaris berdasarkan jumlah penjualan mengungkapkan dominasi lini produk Colorfit, khususnya produk bibir dan alas bedak. Produk seperti Colorfit Velvet Matte Lip Mousse dan Matte Lip Cream masuk ke dalam Klaster 4, yang mengindikasikan kategori unggulan, sedangkan sebagian besar lainnya tergolong dalam Klaster 3, yang merepresentasikan produk stabil dengan potensi untuk dikembangkan lebih lanjut.

4. KESIMPULAN

Berdasarkan hasil segmentasi dan analisis penjualan, dapat disimpulkan bahwa metode K-Means Clustering efektif dalam mengelompokkan produk berdasarkan performa pasar. Penentuan lima klaster sebagai jumlah optimal menghasilkan segmentasi yang mencerminkan karakteristik harga, volume penjualan, dan jumlah varian dari masing-masing produk. Produk-produk dalam Klaster 4 terbukti sebagai kontributor utama terhadap penjualan dan patut dijadikan fokus dalam strategi pengembangan dan promosi. Selain itu, produk-produk dalam Klaster 3 juga menunjukkan potensi tinggi yang dapat ditingkatkan dengan pendekatan pemasaran yang tepat. Penelitian ini memberikan wawasan berharga untuk pengambilan keputusan manajerial dalam pengelolaan portofolio produk dan strategi distribusi.

DAFTAR PUSTAKA

[1] M. Damayanti and A. Setiawan, "Faktor-Faktor Pengaruh Keputusan Pembelian Kosmetik Wardah Generasi Milenial Di Kabupaten Semarang," *El-Iqtishod J. Ekon. Syariah*, vol. 7, no. 2, pp. 1–18, 2023, doi: 10.70136/eliqtishod.v7i2.414.

[2] E. M. Putri, P. B. Wahyudi, S. A. Zahro, S. W. Kirana, R. Hidayat, and M. Ikaningtyas, "Meningkatkan Citra Merek Untuk Memperoleh Keunggulan Bersaing Melalui Perencanaan dan Pengembangan Bisnis Yang Efektif pada Produk Kosmetik," *Econ. Bus. Manag. J.*, vol. 3, no. 01, pp. 134–142, 2024.

[3] N. Kastiawan, B. Huda, E. Novalia, and F. Nurapriani, "Klasterisasi Data Obat dengan Algoritma K-Means (Kasus pada UPTD Puskesmas Curug)," *J. Sains Komput. Inform. (J-SAKTI*, vol. 8, no. 1, pp. 120–130, 2024.

[4] V. Holý, O. Sokol, and M. Černý, "Clustering retail products based on customer behaviour," *Appl. Soft Comput. J.*, vol. 60, 2017, doi: 10.1016/j.asoc.2017.02.004.

[5] D. S. Annam, A. Hananto, F. Nurapriani, and T.

This work is licensed under a Creative Commons Attribution 4.0 International License

Tukino, "Clustering User Sentiment Transportasi Online Gojek Dan Grab Dengan Metode K-Means," *J. Tika*, vol. 8, no. 2, 2023, doi: 10.51179/tika.v8i2.2165.

[6] A. I. Lailani, N. A. Hasanah, D. P. Meilia, and A. Fauzi, "Strategi Pemasaran Inovatif Untuk Meningkatkan Minat Masyarakat Terhadap Produk Tabungan Di BTN Kantor Cabang Syariah Jember Menulis : Jurnal Penelitian Nusantara," *J. Penelit. Nusant.*, vol. 1, no. 3, pp. 442–448, 2025, doi: https://doi.org/10.59435/menulis.v1i3.137.

[7] A. Lia Hananto *et al.*, "Analysis of Drug Data Mining with Clustering Technique Using K-Means Algorithm," in *Journal of Physics: Conference Series*, 2021. doi: 10.1088/1742-6596/1908/1/012024.

[8] T. Tukino and B. Huda, "PENERAPAN ALGORITMA K-MEANS UNTUK MENDUKUNG KEPUTUSAN DALAM PEMILIHAN TEMA TUGAS AKHIR PADA PRODI SISTEM INFORMASI UNIVERSITAS BUANA PERJUANGAN KARAWANG.," *Techno Xplore J. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 1, 2019, doi: 10.36805/technoxplore.v4i1.542.

[9] B. Bangoria, N. Mankad, and V. Pambhar, "A survey on Efficient Enhanced K-Means Clustering Algorithm," *IJSRD -International J. Sci. Res. Dev.*, vol. 1, no. 9online, 2013.

[10] F. Marisa, A. R. Wardhani, W. Purnomowati, A. V. Vitianingsih, A. L. Maukar, and E. W. Puspitarini, "POTENTIAL CUSTOMER ANALYSIS USING K-MEANS WITH ELBOW METHOD," *JIKO (Jurnal Inform. dan Komputer)*, vol. 7, no. 2, 2023, doi: 10.26798/jiko.v7i2.911.

[11] S. Sugiyono and M. Ruswandi, "Pemodelan Pengolahan Citra untuk Klasifikasi Jenis Buah Pisang Menggunakan Metode KNN," *J. Pendidik. dan Konseling*, vol. 4, no. 5, pp. 823–833, 2022.

[12] N. C. Ramadhan, H. H. H, T. Rohana, and A. M. Siregar, "Optimasi Algoritma Machine Learning Menggunakan Seleksi Fitur Xgboost Untuk Klasifikasi Kanker Payudara," *TIN Terap. Inform. Nusant.*, vol. 5, no. 2, pp. 162–171, 2024, doi: 10.47065/tin.v5i2.5408.

[13] F. Rachmawati, J. Jaenudin, N. B. Ginting, and P. Laksono, "Machine Learning for the Model Prediction of Final Semester Assessment (FSA) using the Multiple Linear Regression Method," *J. Tek. Inform.*, vol. 17, no. 1, pp. 1–9, 2024, doi: 10.15408/jti.v17i1.28652.

[14] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," in *IOP Conference Series: Materials Science and Engineering*, 2018. doi: 10.1088/1757-899X/336/1/012017.

[15] Y. Sopyan, A. D. Lesmana, and C. Juliane, "Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, 2022, doi: 10.47065/bits.v4i3.2697.

[16] I. F. Ashari, E. Dwi Nugroho, R. Baraku, I. Novri Yanda, and R. Liwardana, "Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta," *J. Appl. Informatics Comput.*, vol. 7, no. 1, 2023, doi: 10.30871/jaic.v7i1.4947.

[17] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 4th ed. Burlington: Morgan Kaufmann, 2022. doi: 10.1016/C2009-0-61819-5.

[18] M. B. Hendrawan and G. Gunadi, "Segmentasi Layanan Jasa Kantor Akuntan Publik Menggunakan Algoritma K-Means Clustering," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 7, no. 2, pp. 268–276, 2025.

This work is licensed under a Creative Commons Attribution 4.0 International License