Segmentasi Produk Parfum Lokal Dengan K-Means, Python, Dan Visualisasi Interaktif

Noviana Riza[1)], Iga Putri Ramadhani[2)], Dini Aprilia Cahyani[3)*], Kiki Mustaqim[4)]
[1,2,3,4)] Universitas Logistik dan Bisnis Internasional (ULBI)

Received: 25 September 2025 Accepted: 22 December 2025 Published: 28 December 2025

*apriliadini206@gmail.com

Kata Kunci: Segmentasi, parfum local, followers, k-means, visualisasi interaktif.

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Produk parfum brand lokal Indonesia saat ini semakin diminati oleh masyarakat. Keragaman kriteria pada produk parfum memiliki peranan yang sangat penting, seperti tahun berdirinya brand parfum, konsentrasi aroma, ukuran botol, serta harga berdasarkan ukuran, yang dapat membantu konsumen dalam memilih produk parfum brand lokal sesuai dengan preferensi mereka. Selain itu, pelaku usaha parfum juga memerlukan pemahaman yang tepat dalam pengelompokan produk guna meningkatkan kualitas dan daya saing produk yang ditawarkan. Penelitian bertujuan melakukan segmentasi produk parfum brand lokal Indonesia berdasarkan data jumlah pengikut di platform media sosial serta karakteristik setiap klaster, sehingga dapat membantu pelaku usaha dan konsumen dalam memahami produk dan mengambil keputusan yang tepat. Metode yang digunakan adalah K-Means Clustering. Hasil penelitian menunjukkan bahwa 154 produk parfum memiliki keragaman karakteristik yang signifikan. Secara statistik, rata-rata jumlah pengikut adalah 22.625, tahun berdiri rata-rata 2021, ukuran botol rata-rata 50,46 ml, serta harga rata-rata per ml sebesar Rp4.810. Jenis konsentrasi parfum didominasi oleh Eau de Parfum (EDP) sebesar 84,4%. Proses segmentasi menghasilkan empat klaster, yaitu Klaster 0 (populer) dengan jumlah pengikut tertinggi, Klaster 1 (newcomers) yang mencakup mayoritas data, Klaster 2 (medium) dengan ukuran botol terbesar dan harga per ml termurah, serta Klaster 3 (luxury) dengan harga per ml tertinggi. Visualisasi interaktif menggunakan Looker Studio dilakukan menggunakan Looker Studio untuk menyajikan hasil analisis secara informatif dan mudah dipahami

PENDAHULUAN

Pesatnya pertumbuhan industri parfum di Indonesia membuka peluang bisnis besar sekaligus meningkatkan persaingan antar brand lokal yang menawarkan produk berkualitas dengan harga terjangkau[1]. Parfum kini bukan sekadar produk wewangian, melainkan instrumen penting untuk meningkatkan kepercayaan diri dan menciptakan impresi positif[2]. Namun, keberagaman varian dari sisi konsentrasi, ukuran, hingga komposisi aroma yang unik justru menimbulkan kompleksitas bagi konsumen dalam memilih produk yang sesuai dengan preferensi mereka[3]. Di sisi lain, perusahaan parfum seringkali menghadapi tantangan dalam mengolah data penjualan yang besar untuk menemukan pola yang relevan guna meningkatkan kinerja bisnis[4]. Meskipun data tersedia, strategi pemasaran dan rekomendasi produk sering kali belum optimal karena kurangnya pemahaman mendalam mengenai pengelompokan karakteristik produk. Oleh karena itu, diperlukan sebuah pendekatan untuk mengklasifikasikan varian parfum yang beragam agar brand dapat merancang strategi pemasaran yang lebih efektif dan personal[5].

Penelitian bertujuan mengatasi permasalahan tersebut melalui penerapan Data Mining dengan metode K-Means Clustering. Algoritma dipilih karena efisiensinya dalam mengelompokkan data berdasarkan kemiripan karakteristik secara objektif [6]. Dengan melakukan segmentasi produk parfum brand lokal ke dalam klaster-klaster tertentu, pelaku usaha dapat memetakan karakteristik produknya secara lebih spesifik. Untuk memperkuat penyampaian hasil analisis, penelitian ini juga mengintegrasikan

visualisasi interaktif melalui Looker Studio. Hal ini bertujuan agar hasil segmentasi tidak hanya akurat secara teknis, tetapi juga mudah dipahami oleh pelaku bisnis sebagai dasar pengambilan keputusan strategis yang lebih bijak.

TINJAUAN LITERATUR

Secara konseptual, produk merupakan pemahaman subjektif produsen terhadap sesuatu yang ditawarkan untuk memenuhi kebutuhan konsumen dengan menyesuaikan kapasitas dan daya beli pasar (Sari R., 2023). Hal ini menunjukkan bahwa produk bukan sekadar barang, melainkan upaya strategis organisasi dalam menjawab kondisi pasar[7]. Dalam industri parfum, tantangan utama adalah melakukan segmentasi produk yang tepat guna mengarahkan strategi pemasaran secara efektif. Penelitian terdahulu oleh Yansah, Wulansari, dan Alameka(2022) pada FR Parfum berhasil mengidentifikasi segmen produk terlaris hingga kurang diminati berdasarkan data penjualan [8]. Selain itu, Rahayu et al.(2023) juga menegaskan bahwa segmentasi pada toko parfum lokal sangat krusial untuk pemetaan strategi harga, promosi, serta meningkatkan loyalitas pelanggan dan efisiensi operasional[9].

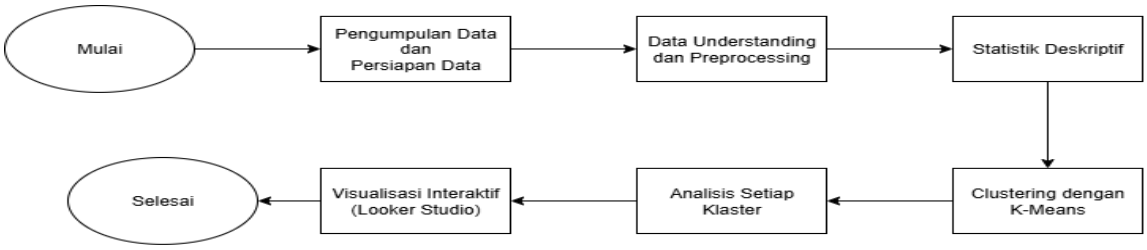
Untuk melakukan segmentasi tersebut, algoritma K-Means Clustering menjadi pendekatan yang dominan karena kemampuannya mengelompokkan data ke dalam klister berdasarkan karakteristik tertentu secara objektif. Harahap, Lubis, dan Situmorang(2021) menyatakan bahwa penerapan K-Means dalam dunia pemasaran memungkinkan perusahaan merancang strategi yang lebih personal dan efisien melalui pengelompokan perilaku pelanggan[10]. Dalam konteks penelitian ini, metode tersebut digunakan untuk mengklasifikasikan produk parfum brand lokal agar setiap klaster dapat dianalisis karakteristik uniknya secara mendalam.

Sebagai elemen pendukung dalam analisis bisnis, visualisasi data berperan mengubah data rumit menjadi format visual yang lebih mudah dipahami untuk menemukan pola tersembunyi[11]. Pemanfaatan tools seperti Tableau telah membuktikan bahwa visualisasi membantu instansi atau perusahaan dalammengambil keputusan setelah data ditransformasikan ke dalam format grafik atau tabel yang informatif[12]. Pentingnya aspek visual ini juga digarisbawahi oleh Husein, Lubis, dan Harahap(2021) yang menekankan penggunaan dashboard interaktif untuk mempercepat akurasi pengambilan keputusan berbasis data[13]

Sejalan dengan teori tersebut, penelitian ini menggunakan Looker Studio untuk menyajikan hasil segmentasi parfum brand lokal secara interaktif. Integrasi antara metode K-Means dan visualisasi yang komunikatif diharapkan dapat memberikan insight yang lebih jelas bagi pelaku usaha dalam merumuskan strategi pemasaran yang efisien serta membantu konsumen dalam proses pemilihan produk yang lebih terinformasi.

METODE PENELITIAN

Metode penelitian berfokus pada proses analisis segmentasi produk parfum brand lokal Indonesia berdasarkan karakteristik produknya, dengan memanfaatkan Metode K-Means Clustering serta visualisasi interaktif melalui Looker Studio. Berikut tahapan yang terdapat pada diagram diabawah ini.



Gambar 1. Alur Metode Penelitian

Penjelasan tahapan utama dalam proses analisis meliputi:

1. Pengumpulan Data dan Persiapan Data
   Penelitian mengandalkan dataset berbentuk CSV yang diunduh dari situs Kaggle, berjudul "Indonesian Local Perfume Brands", yang menjadi acuan utama dalam pelaksanaan penelitian yang bertajuk "Analisis Segmentasi Produk Parfum Brand Lokal Indonesia Berdasarkan Karakteristik Produk dengan Metode K-Means Clustering dan Visualisasi Interaktif". Dataset awal memiliki deskripsi tabel yang cukup ekstensif, mencakup 154 baris dan 38 kolom. Sebelum masuk ke tahapan analisis, dataset

yang diperoleh harus melalui proses pembersihan data (data cleaning) yang ketat untuk memastikan kualitas data. Proses ini, termasuk membersihkan nilai yang kosong (missing values) dan penanganan anomali data lainnya, dilakukan secara terstruktur menggunakan pustaka Pandas dalam lingkungan pemrograman Python, yang dieksekusi melalui platform Google Colab atau VSCode. Tahapan persiapan data ini krusial untuk menjamin bahwa data yang digunakan memenuhi standar kualitas yang dibutuhkan untuk pemodelan K-Means Clustering.

2. Data Understanding dan Preprocessing

   Tahapan Data Understanding dan Preprocessing merupakan langkah fondasi yang krusial sebelum analisis data inti dilakukan. Data Understanding berfokus pada pemahaman mendalam terhadap struktur data yang digunakan, yang meliputi pengecekan dimensi dataset (jumlah baris dan kolom), identifikasi tipe data setiap variabel untuk menjamin konsistensi, hingga deteksi nilai yang kosong (missing values) dan pengecekan nilai unik pada setiap kolom. Proses ini esensial untuk mengidentifikasi potensi masalah dan asumsi distribusi data yang diperlukan untuk klustering. Selanjutnya, Preprocessing Data adalah proses yang vital untuk memastikan data yang akan diolah memenuhi standar kualitas dan asumsi distribusi yang dibutuhkan, terutama dalam konteks pengelompokan menggunakan K-Means [14]. Tahapan ini berfokus pada pembersihan (cleaning), perapian, dan transformasi data—seperti penanganan outliers atau encoding variabel kategorikal—sebelum memasuki analisis. Keseluruhan tahapan preprocessing ini sangat penting untuk mencegah bias dan memastikan hasil klustering yang akurat dan valid.

   a. Standarisasi

      Standarisasi data numerik agar skala antar fitur seragam. Kolom numerik yang akan dinormalisasi atau *scaling* seperti kolom *followers, year, size*, dan *price_per_ml*. Normalisasi menjamin variabel-variabel dengan skala atau rentang nilai yang berbeda dapat diukur dan dibandingkan secara setara [15]. Dalam rangka mempersiapkan data numerik untuk algoritma K-Means Clustering, tahap standarisasi merupakan langkah yang wajib dilakukan. Proses ini secara khusus diterapkan dengan menggunakan metode StandardScaler. Penggunaan StandardScaler bertujuan untuk mentransformasi data numerik sehingga setiap fitur memiliki nilai rata-rata nol dan standar deviasi satu. Hal ini menjadi krusial karena algoritma K-Means sangat sensitif terhadap perbedaan skala fitur. Apabila data tidak distandarisasi, fitur dengan rentang nilai yang lebih besar, seperti jumlah followers, cenderung mendominasi perhitungan jarak Euclidean antar data, sehingga mengurangi pengaruh fitur lain yang juga memiliki tingkat kepentingan yang sama, seperti harga per ml. Dengan dilakukannya standarisasi, skala antar fitur menjadi sebanding, sehingga setiap variabel dapat memberikan kontribusi yang setara dan tidak saling mendominasi dalam proses pembentukan klaster. Berikut ini merupakan rumus standarisasi data.

$$X_{stand} = \frac{X - mean(X)}{standard\ deviation\ (X)}$$

      Keterangan:

      1. X yaitu nilai asli data
      2. mean(X) yaitu average fitur
      3. standard deviation(X) yaitu standar deviasi fitur
      4. X_stand yaitu hasil standarisasi

   b. Encoding

      Tahapan selanjutnya dalam preprocessing adalah Encoding, yang memiliki tujuan vital untuk mengubah variabel yang bersifat kategorikal atau tekstual menjadi format data numerik. K-Means Clustering adalah algoritma yang berbasis perhitungan jarak (distance-based), sehingga ia hanya dapat memproses fitur-fitur yang direpresentasikan dalam bentuk angka. Untuk mencapai tujuan ini, penelitian ini menggunakan metode One-Hot Encoding (OHE). Secara spesifik pada dataset Indonesian Local Perfume Brands, kolom concentration—yang berisi kategori seperti Eau de Parfum (EDP) atau Eau de Toilette (EDT)—merupakan salah satu kolom kategorial utama yang harus disertakan dalam proses clustering. Oleh karena itu, penerapan One-Hot Encoding sangat diperlukan untuk mengubah setiap kategori unik

menjadi kolom biner (0 atau 1) yang bersifat numerik, sehingga data siap untuk diolah dan dianalisis secara efektif oleh algoritma K-Means.

3. Statistik Deskriptif

Tahapan Statistik Deskriptif merupakan langkah dasar dalam metodologi penelitian. Pelaksanaan tahap ini bertujuan untuk mendapatkan pemahaman awal yang menyeluruh mengenai karakteristik umum dan distribusi dari seluruh dataset yang telah melalui proses preprocessing. Dengan menganalisis ringkasan numerik seperti rata-rata, median, rentang, dan standar deviasi untuk setiap variabel, peneliti dapat mengidentifikasi pola, kecenderungan, dan potensi anomali (outliers) dalam data sebelum masuk ke proses pengolahan yang lebih kompleks. Pemahaman mendalam ini sangat penting karena menjadi dasar yang penting dan referensi awal sebelum data dimasukkan ke dalam proses segmentasi atau klastering, membantu dalam memvalidasi hasil segmentasi yang terbentuk di akhir analisis.

	followers	year	size_ml	price	price_per_ml
count	154.000000	154.000000	154.000000	154.000000	154.000000
mean	22625.584416	2021.590909	50.461039	229874.188312	4810.010455
std	63342.029385	1.740008	20.513070	166161.745789	3179.502659
min	9.000000	2013.000000	10.000000	34900.000000	1076.000000
25%	1087.250000	2020.250000	35.000000	143575.000000	3069.167500
50%	3350.500000	2022.000000	50.000000	185000.000000	3980.000000
75%	15500.000000	2023.000000	50.000000	249000.000000	5144.167500
max	541000.000000	2024.000000	105.000000	995000.000000	19900.000000

Gambar 2. Statistika Deskriptif pada Colab

Pada gambar 2 didapatkan sebuah hasil analisis, terlihat bahwa rata-rata jumlah followers brand parfum lokal berada di angka 22.625, menunjukkan variasi yang cukup besar antara brand kecil dan brand yang sudah lebih dikenal. Rata-rata tahun berdirinya brand berada di 2021, yang berarti sebagian besar brand parfum lokal yang ada dalam dataset termasuk pemain baru dalam industri. Ukuran botol parfum juga cukup konsisten, dengan rata-rata 50,46 ml, yang sesuai dengan ukuran standar produk parfum. Dari sisi harga, rata-rata harga produk adalah sekitar Rp 229.874, sedangkan harga per ml berada di angka Rp 4.810. Nilai-nilai ini membantu memberikan gambaran awal tentang pola harga dan strategi pemasaran yang digunakan brand parfum lokal. Tahapan ini menjadi dasar penting sebelum melakukan clustering, karena memberikan gambaran awal mengenai sebaran data dan karakteristik utama tiap variabel.

4. *Clustering* dengan *K-Means*

K-Means Clustering merupakan teknik non-hirarki untuk mengelompokkan data, dengan menempatkan data yang memiliki fitur serupa dalam satu klaster dan memisahkan data yang memiliki karakteristik berbeda ke dalam klaster lain [16] . Dalam kerangka kerja analisis segmentasi ini, metode K-Means Clustering berfungsi sebagai algoritma fundamental yang digunakan untuk mengelompokkan data produk parfum berdasarkan kemiripan karakteristik multidimensi. Mengingat sensitivitas K-Means terhadap inisialisasi dan jumlah klaster, penentuan nilai k (jumlah klaster) yang paling tepat merupakan tahapan krusial untuk menjamin validitas dan optimalitas hasil pemodelan. Berikut merupakan rumus jarak Euclidean yang digunakan dalam algoritma K-Means dalam menghitung jarak antara setiap data ke centroid.

$$d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \ldots + (x_{ki} - x_{kj})^2}$$

Keterangan:

1. d_{ij} yaitu kuadrat jarak *Euclidean* antara data ke-1 dan centroid ke-j
2. x_{ki} yaitu nilai fitur ke-k dari data ke-i
3. x_{kj} yaitu nilai fitur ke-k dari centorid ke-j

Rumus berikut menghitung centroid baru pada setiap iterasi K-Means, dengan cara mengambil rata-

rata dari semua data yang termasuk dalam klaster. Proses ini akan terus dilakukan secara berulang hingga posisi centroid stabil.

$$C_j = \frac{1}{|S_j|} \sum_{X_i \epsilon S_j} x_i$$

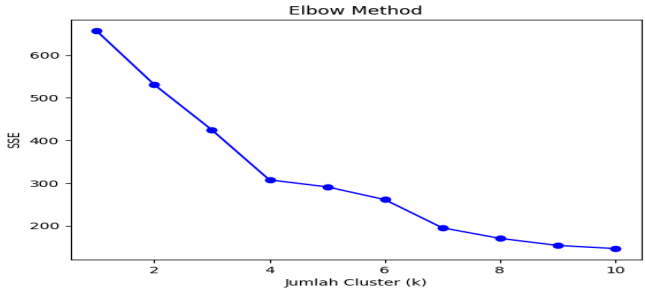
Keterangan:

1. S_j yaitu kumpulan data dalam klaster ke-j
2. C_j yaitu centorid atau pusat klaster ke-j
3. $X_i$yaitu titik data yang termasuk dalam S_j

Penentuan nilai k tidak hanya didasarkan pada satu metrik, melainkan melalui evaluasi silang (*cross-validation*) dari dua teknik validasi klaster yang paling teruji yaitu Metode Elbow dan Silhouette Score. Kedua metode secara konsisten menghasilkan hasil yang sama, yaitu bahwa jumlah klaster terbaik adalah empat (k=4). Metode Elbow, yang memplot nilai *Sum of Squared Errors* (SSE), secara visual mengindikasikan bahwa pada titik k=4, terjadi penurunan laju SSE yang mulai melambat secara signifikan, membentuk elbow atau 'siku' yang menandakan bahwa penambahan klaster lebih lanjut tidak lagi memberikan gain substansial dalam mengurangi variabilitas intra-klaster. Temuan ini diperkuat secara statistik oleh Silhouette Score, di mana nilai koefisien tertinggi diperoleh tepat pada k=4. Nilai Silhouette yang tinggi pada k=4 ini secara tegas menegaskan bahwa pada pembagian empat segmen

Produk parfum brand lokal Indonesia saat ini semakin diminati oleh masyarakat. Keragaman kriteria pada produk parfum memiliki peranan yang sangat penting, seperti tahun berdirinya brand parfum, konsentrasi aroma, ukuran botol, serta harga berdasarkan ukuran, yang dapat membantu konsumen dalam memilih produk parfum brand lokal sesuai dengan preferensi mereka. Selain itu, pelaku usaha parfum juga memerlukan pemahaman yang tepat dalam pengelompokan produk guna meningkatkan kualitas dan daya saing produk yang ditawarkan.

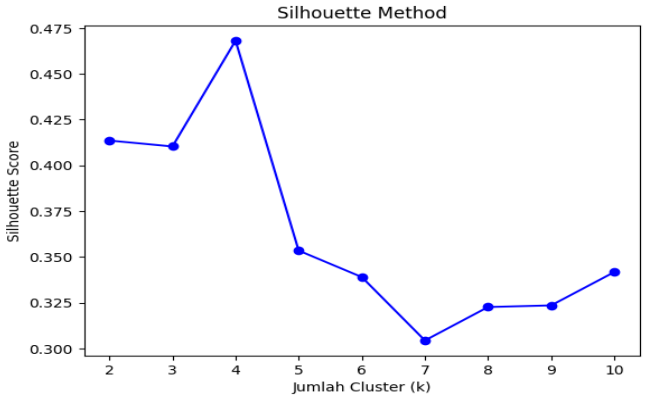
Penelitian bertujuan melakukan segmentasi produk parfum brand lokal Indonesia berdasarkan data jumlah pengikut di platform media sosial serta karakteristik setiap klaster, sehingga dapat membantu pelaku usaha dan konsumen dalam memahami produk dan mengambil keputusan yang tepat. Metode yang digunakan dalam penelitian ini adalah K-Means Clustering. Hasil penelitian menunjukkan bahwa 154 produk parfum memiliki keragaman karakteristik yang signifikan. Secara statistik, rata-rata jumlah pengikut adalah 22.625, tahun berdiri rata-rata 2021, ukuran botol rata-rata 50,46 ml, serta harga rata-rata per ml sebesar Rp4.810. Jenis konsentrasi parfum didominasi oleh Eau de Parfum (EDP) sebesar 84,4%. Proses segmentasi menghasilkan empat klaster, yaitu Klaster 0 (populer) dengan jumlah pengikut tertinggi, Klaster 1 (newcomers) yang mencakup mayoritas data, Klaster 2 (medium) dengan ukuran botol terbesar dan harga per ml termurah, serta Klaster 3 (luxury) dengan harga per ml tertinggi. Visualisasi interaktif dilakukan menggunakan Looker Studio untuk menyajikan hasil analisis secara informatif dan mudah dipahami.oduntuk analisis segmentasi pasar produk parfum selanjutnya.



Gambar 3. Metode Elbow dan Silhouette Score

Pada Gambar 3, disajikan hasil visualisasi dari Metode Elbow yang digunakan secara spesifik untuk menentukan jumlah klaster optimal (k) dalam algoritma K-Means. Grafik tersebut secara eksplisit memperlihatkan tren penurunan nilai SSE (Sum of Squared Errors) yang berkelanjutan seiring dengan penambahan jumlah klaster. Titik klaster keempat (k=4) merupakan titik yang paling krusial, karena pada posisi tersebut, grafik menunjukkan adanya bentuk siku yang jelas (elbow). Titik siku ini mengindikasikan adanya perubahan signifikan di mana laju penurunan nilai SSE mulai melambat

drastis. Fenomena ini secara metodologis menunjukkan bahwa penambahan klaster setelah k=4 tidak lagi memberikan gain atau peningkatan signifikansi dalam mengurangi variabilitas dalam klaster (intra-cluster variance) yang cukup berarti. Oleh karena itu, k=4 ditetapkan sebagai jumlah klaster yang paling optimal dan efisien. Dengan nilai SSE yang diperoleh di klaster keempat berada di sekitar 300, nilai ini kemudian dijadikan sebagai acuan dasar untuk proses klustering dan analisis segmentasi produk parfum selanjutnya.



Gambar 4. Metode Elbow dan Silhouette Score

Pada Gambar 4 ditampilkan grafik metode Silhouette yang digunakan untuk mengevaluasi kualitas pemisahan antar klaster. Nilai Silhouette Score merepresentasikan seberapa baik objek berada dalam klasternya masing-masing—semakin mendekati 1, semakin baik kualitas pengelompokkannya. Berdasarkan grafik tersebut, nilai Silhouette Score tertinggi terlihat pada klaster ke-4, yang menunjukkan bahwa pemilihan jumlah klaster sebanyak empat memberikan pemisahan yang paling optimal dibandingkan jumlah klaster lainnya. Hal ini menguatkan hasil dari metode Elbow sebelumnya, sehingga nilai k = 4 dapat dianggap sebagai jumlah klaster terbaik untuk data parfum brand lokal ini.

5. Analisis Klaster

   Tahapan ini menganalisis hasil dari *clustering*, dimana menyusun ringkasan setiap klaster. Tahapan ini bertujuan agar memberikan makna dari setiap klaster dan mengidentifikasi suatu kriteria utama atau perbedaan utama. Misalkan terdapat 4 klaster yaitu klaster 0 merupakan *populer* yang berarti pembeli sudah mengenal *brand* lokal populer, lalu klaster 1 *newcomers* yaitu cocok untuk pembeli yang baru mulai memakai parfum *brand* lokal, klaster 2 merupakan *medium*, cocok bagi pengguna parfum aktif, dan terakhir klaster 3 cocok untuk pecinta parfum *luxury*. Jadi analisis ini membantu memberikan *insight* atau wawasan baru. Selain itu, analisis ini membantu pengguna parfum dalam memilih parfum sesuai preferensi mereka, dan juga membantu pelaku usaha untuk mendapatkan pemahaman lebih dalam sehingga dapat meningkatkan kualitas produk *brand* parfum mereka menjadi lebih baik.

6. Visualisasi

   Pada tahapan visualisasi hasil penelitian, peneliti mengambil langkah strategis dengan memanfaatkan platform Looker Studio. Pemilihan Looker Studio ini didasarkan pada kemampuannya yang unggul dalam menyediakan *dashboard* visualisasi data interaktif secara *real-time*. Tujuannya utama adalah untuk menginterpretasikan hasil analisis *clustering* yang telah dilakukan. Dengan sifat interaktifnya, *dashboard* ini memungkinkan untuk melakukan eksplorasi data seperti memfilter, menelusuri detail, dan membandingkan hasil antar klaster secara dinamis. Hal ini secara signifikan meningkatkan kualitas penyajian data menjadi lebih informatif, transparan, dan menarik.

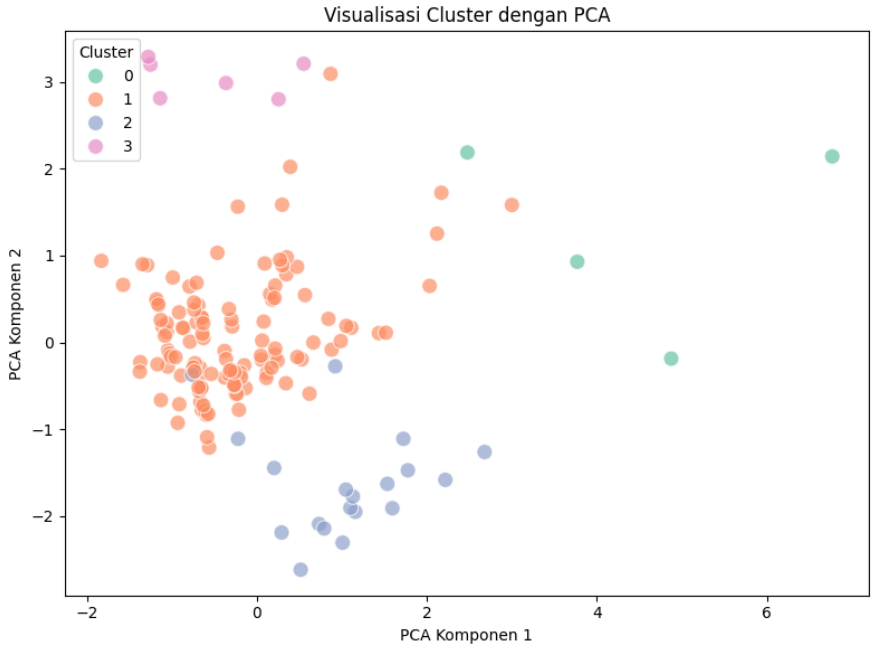
   Data mentah hasil *clustering* yang divisualisasikan mencakup rangkuman karakteristik setiap klaster. Visualisasi yang disajikan meliputi grafik distribusi seperti *bar chart* untuk menganalisis penyebaran data kuantitatif seperti jumlah pengikut yaitu followers per klaster, volume produk yaitu size per klaster, dan harga per mililiter ysitu harga/ml per klaster. Selain itu, digunakan pula Pie Chart untuk menunjukkan proporsi data kategori, seperti komposisi konsentrasi ketahanan parfum per klaster misalnya Eau de Parfum, Eau de Toilette dan jumlah persentase brand produk per klaster. Kemudian untuk memberikan *insight* statistik mendalam, disediakan tabel statistik deskriptif yang menyajikan ringkasan numerik (seperti rata-rata, median, standar deviasi, nilai minimum, dan maksimum untuk

setiap klaster dan setiap variabel yang digunakan, sehingga karakteristik masing - masing klaster dapat teridentifikasi dengan akurat. Kombinasi visualisasi dan tabel deskriptif ini memastikan bahwa temuan dari analisis *clustering* dapat dikomunikasikan secara efektif dan mudah dipahami.

HASIL PENELITIAN

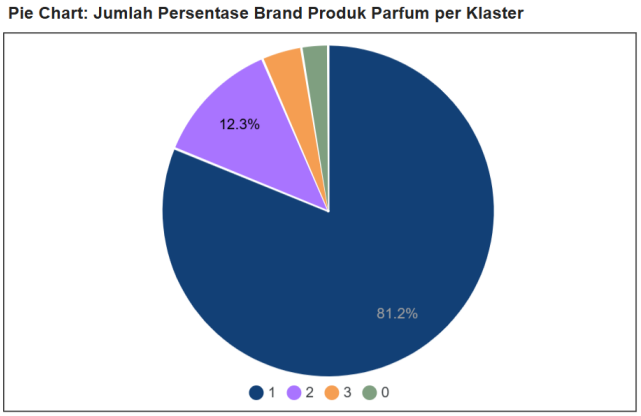
Dari analisis deskriptif terhadap total 154 data produk parfum dari merek-merek lokal (brandlokal), diperoleh beberapa gambaran kunci mengenai karakteristik pasar. Secara umum, pasar ini didominasi oleh pemain-pemain baru, yang tercermin dari rata-rata tahun berdiri merek yang berada di sekitar 2021. Hal ini mengindikasikan bahwa industri parfum lokal berada dalam fase pertumbuhan yang dinamis dan kompetitif (nascent market). Karakteristik produk menunjukkan preferensi yang kuat pada konsentrasi Eau de Parfum (EDP), yang merupakan mayoritas signifikan, mencapai 84,4% dari total produk, menegaskan kecenderungan konsumen terhadap ketahanan aroma yang lebih lama. Sementara itu, ukuran botol yang paling dominan di pasar adalah 50 ml, menjadikannya standar kemasan yang paling disukai. Aspek popularitas merek menunjukkan heterogenitas yang ekstrem, di mana jumlah pengikut media sosial merek-merek ini bervariasi sangat luas, mulai dari 9 pengikut hingga mencapai angka fantastis 541.000. Rentang variasi yang besar ini mencerminkan perbedaan yang substansial dalam tingkat popularitas, jangkauan pasar, dan strategi digital marketing antar merek. Untuk menggali pola dan memisahkan heterogenitas pasar ini menjadi segmen yang lebih homogen, penelitian ini kemudian menerapkan metode K-Means Clustering, yang menghasilkan 4 klaster optimal untuk analisis lebih lanjut:

1. Klaster 0 *(Populer)*: *Brand* dengan *followers* tertinggi (rata-rata 362.250).
2. Klaster 1 *(Newcomers)*: Mencakup 81,2% dari total produk, kriteria umum dan seimbang.
3. Klaster 2 *(Medium)*: Ukuran botol terbesar (97,37 ml) dan harga per ml terendah (Rp 3.283).
4. Klaster 3 (*Luxury*): Harga per ml paling tinggi (Rp 17.526).



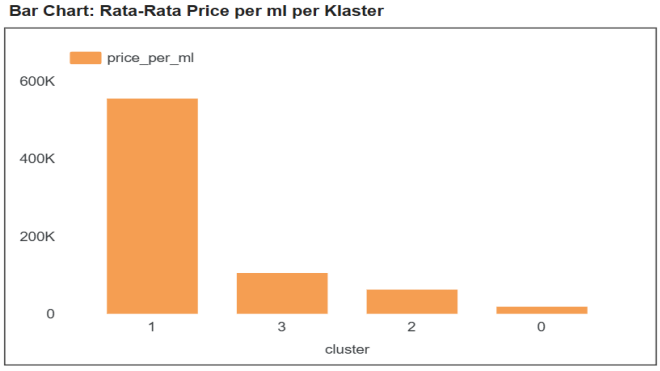
Gambar 5 Visualisasi Klaster dengan PCA

Pada gambar 5 menunjukkan visualisasi hasil clustering menggunakan PCA (*Principal Component Analysis*). PCA digunakan untuk mereduksi dimensi data sehingga setiap klaster dapat ditampilkan dalam grafik dua dimensi, tanpa mengurangi pola utama dalam datanya. Pada visualisasi tersebut terlihat bahwa data terbagi menjadi empat klaster yang memiliki sebaran dan posisi berbeda. Setiap warna mewakili satu klaster yang terbentuk dari algoritma K-Means. Klaster-klaster tersebut menunjukkan bahwa produk parfum lokal memiliki karakteristik yang cukup berbeda satu sama lain, sehingga dapat dipisahkan secara visual melalui dua komponen utama PCA. Perbedaan posisi antar klaster mengindikasikan variasi karakteristik seperti harga, ukuran botol, tahun brand, hingga jumlah followers. Sementara itu, titik-titik yang berada berdekatan menunjukkan produk dengan karakteristik yang mirip, sedangkan titik yang lebih jauh menandakan produk dengan perbedaan karakteristik yang lebih besar.



Gambar 6. Visualisasi dari Looker Studio Pie Chart

Pada Gambar 6 ditampilkan visualisasi berbentuk Pie Chart dari Looker Studio yang menunjukkan persentase jumlah brand parfum pada setiap klaster. Dari grafik tersebut terlihat bahwa klaster 1 mendominasi, dengan jumlah brand mencapai 81.2%, yang berarti sebagian besar produk parfum lokal memiliki karakteristik yang mirip dan tergolong dalam satu kelompok besar tersebut. Selanjutnya, klaster 2 menyumbang 12.3%, menunjukkan adanya kelompok produk yang memiliki ciri khas berbeda namun jumlahnya tidak sebanyak klaster utama. Kemudian disusul oleh klaster 3, serta klaster 0 yang memiliki porsi paling kecil. Visualisasi ini membantu memperlihatkan seberapa besar distribusi tiap klaster, sekaligus memberikan gambaran bahwa struktur data parfum lokal cenderung terkonsentrasi pada satu klaster dominan, sedangkan klaster lainnya mewakili segmen yang lebih spesifik dan jumlahnya lebih sedikit.



Gambar 7. Visualisasi dari Looker Studio Bar Chart

Berdasarkan visualisasi Bar Chart Rata-Rata Harga per ml per Klaster, terlihat adanya perbedaan harga yang ekstrem antar segmen pasar, di mana Klaster 1 sebagai kategori ultra-premium dengan rata-rata harga per ml yang mendominasi, mencapai kisaran 560.000. Angka ini menunjukkan produk di klaster tersebut memiliki nilai yang jauh lebih tinggi dan signifikan dibandingkan klaster lainnya, yang mungkin didorong oleh kualitas eksklusif atau brand value. Di bawah Klaster 1, terdapat Klaster 3 yang menunjukkan rata-rata harga yang jauh lebih rendah di kisaran 110.000, diikuti oleh Klaster 2 dengan rata-rata harga sekitar 70.000, menempatkan kedua klaster ini pada segmen harga menengah. Terakhir, Klaster 0 mewakili segmen ekonomis dengan rata-rata harga per ml paling rendah, hanya sekitar 20.000. Dengan demikian, proses klustering ini berhasil memisahkan produk secara tajam berdasarkan tingkat harga, dari segmen ultra-premium hingga segmen yang paling sensitif terhadap harga.

	cluster	brand	followers	year	size_ml	price_per_ml
1.	3	6	6,918.83	2020.5	50	17,526.67
2.	2	19	20,084.79	2021.421052631579	97.37	3,283.09
3.	1	125	12,897.73	2021.736	42.97	4,439.17
4.	0	4	362,250	2019.5	62.5	4,576.74

1 - 4 / 4

Gambar 8 Visualisasi Tabel dari Looker Studio Statistik per Klaster

Gambar 8 merupakan tabel ringkasan rata-rata fitur per klaster yang menunjukkan segmentasi yang jelas berdasarkan karakteristik produk, popularitas, dan harga. Klaster 0 menonjol sebagai segmen dengan

rata-rata harga per ml paling tinggi di 4.576,74 dan memiliki followers rata-rata yang fantastis, mencapai lebih dari 362 ribu, meskipun hanya terdiri dari 4 brand saja, yang mengindikasikan segmen ini diisi oleh merek-merek yang sangat populer dan berharga premium. Berbeda dengan Klaster 0, Klaster 1 justru dicirikan oleh jumlah rata-rata brand yang paling banyak, yaitu 125 brand, meskipun rata-rata ukuran produknya menjadi yang terkecil yaitu 42,97 ml dengan rata-rata harga di kisaran 4.439,17. Sementara itu, Klaster 2 menunjukkan rata-rata followers yang tinggi yaitu 20.084,79 dari 19 brand dengan ukuran produk yang besar 97,37 ml, namun memiliki rata-rata harga per ml yang terendah kedua 3.283,09. Terakhir, Klaster 3 menjadi segmen dengan rata-rata harga per ml paling minimal, yaitu 17.526,67, ditandai dengan rata-rata ukuran produk yang kecil yaitu 50 ml dan rata-rata followers yang cukup rendah. Secara keseluruhan, data ini merefleksikan bahwa klustering berhasil memisahkan segmen berdasarkan kombinasi nilai produk, jangkauan merek, dan popularitasnya di media sosial.Pembahasan

Hasil *clustering* menunjukkan adanya segmentasi yang jelas pada produk parfum *brand* lokal. Klaster 0 mengindikasikan bahwa *brand* populer memiliki basis pengikut kuat dan kemungkinan lebih dikenal luas. Klaster 1 mendominasi dengan proporsi produk terbesar, kemungkinan mencerminkan *brand* baru yang sedang tumbuh. Klaster 2 berfokus pada efisiensi seperti ukuran besar, harga murah per ml, sementara Klaster 3 menargetkan segmen eksklusif yang menghargai kualitas premium. Visualisasi interaktif di Looker Studio memperkuat pemahaman terhadap hasil *clustering* dengan tampilan yang informatif. Filter interaktif, tabel ringkasan statistik, dan grafik yang membandingkan masing-masing klaster secara visual, memberikan nilai tambah dalam mendukung analisis dan pengambilan keputusan.

KESIMPULAN

Penelitian berhasil melakukan segmentasi pasar produk parfum brand lokal Indonesia secara efektif dengan algoritma K-Means Clustering menggunakan pemrograman python dan visualisasi interaktif Looker Studio terhadap 154 produk, parfum lokal menunjukkan karakteristik rata-rata merek berdiri tahun 2021, didominasi oleh konsentrasi EDP (84,4%), dan mengadopsi ukuran standar 50 ml. Heterogenitasnya ekstrem pada popularitas, di mana jumlah followers media sosial bervariasi luas antara 9 hingga 541.000, yang menjadi justifikasi kuat. Penerapan K-Means secara optimal membagi data menjadi empat klaster yang terdefinisi dengan jelas, merefleksikan pembagian pasar berdasarkan kombinasi nilai, popularitas, dan karakteristik produk: Klaster 0 (Populer) yaitu dimana pembeli sudah mengenal brand lokal populer, memiliki rata-rata followers tertinggi sebanyak 362.250; Klaster 1 (Newcomers) cocok untuk pembeli yang baru mulai memakai parfum lokal, yang mencakup mayoritas produk, yaitu sebanyak 81.2% dari total, dengan karakteristik yang cenderung seimbang; Klaster 2 (Medium) cocok untuk pembeli sebagai pengguna parfum aktif, ditandai dengan rata-rata ukuran produk terbesar 97.37 ml dan rata-rata harga per ml terendah Rp 3.283; Klaster 3(Luxury) cocok untuk pecinta parfum mahal, dengan ciri khas rata-rata harga per ml yang paling tinggi Rp 17.526 dibandingkan klaster lainnya.

Secara keseluruhan, temuan tidak hanya memberikan gambaran menyeluruh tentang struktur pasar parfum lokal, tetapi juga menghasilkan peta jalan strategis bagi pelaku usaha secara real-time dan dapat mendukung pengambilan keputusan yang lebih baik. Pengembangan lebih lanjut dengan segmentasi mendalam memungkinkan produk *brand* lokal mengoptimalkan alokasi sumber daya, menyesuaikan strategi *pricing* dan promosi, serta merancang pengembangan produk (*product formulation*) yang spesifik sesuai kebutuhan dan karakteristik unik dari masing-masing klaster.

REFERENCES

[1] R. Riadi dan Mesran, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Analisa Penjualan Parfume," *Journal of Informatics, Electrical and Electronics Engineering*, vol. 2, no. 4, hlm. 138–145, Jun 2023, doi: 10.47065/jieee.v2i4.1181.

[2] R. D. Mutia dan D. U. Sari, "Analisis Konten Strategi Komunikasi Pemasaran Produk Parfum Bermerek Lokal HMNS Melalui Media Sosial Instagram," *Jurnal Penelitian Inovatif*, vol. 4, no. 1, hlm. 49–58, Feb 2024, doi: 10.54082/jupin.255.

[3] F. Banu, J. Saiyad, dan S. Dedhia, *A Comparative Study on the Perfume Industry and Customer Perception in India and Dubai: An Empirical Analysis, Educational Administration: Theory and Practice*, vol. 2024, no. 5. 2024, hlm. 8229–8235. doi: 10.53555/kuey.v30i5.4332.

[4] F. A. Maresti, W. I. Rahayu, M. B. B. C. Lustin, dan T. H. Pakpahan, *Implementasi K-Means Untuk*

Melakukan Segmentasi Produk Berdasarkan Data Transaksi Retail, vol. 9, no. 1. 2025.

[5] R. Amadea dan W. Roswinanto, *Promotion Strategy Of Middle Segment Perfume In Indonesian Market*. 2023.

[6] B. de C. L. Rodrigues, V. V. Santana, L. de P. Queiroz, C. M. Rebello, dan I. B. R. Nogueira, "Scents of AI: Harnessing Graph Neural Networks to Craft Fragrances Based on Consumer Feedback," 5 Oktober 2023. doi: 10.20944/preprints202310.0247.v1.

[7] S. Rahayu, "Strategi Pemasaran Produk Dalam Meningkatkan Kepuasan Pelanggan," *Jurnal Penelitian Dan Pengkajian Ilmiah Sosial Budaya*, vol. 2, no. 1, hlm. 109–113, Jan 2023, doi: 10.47233/jppisb.v2i1.705.

[8] H. Yansah, T. Tri Wulansari, dan F. Alameka, *Penerapan Algoritma K-Means Dalam Clustering Produk Terlaris Pada Fr Parfum*, vol. 4, no. 2. 2022, hlm. 83.

[9] M. S. Rahayu, I. R. Yunita, dan C. R. A. Widiawati, "Optimization of Perfume Sales through Data Mining with K-Means Algorithm," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 3, hlm. 1715, Jul 2024, doi: 10.30865/mib.v8i3.7922.

[10] M. Harahap, Y. Lubis, dan Z. Situmorang, "Data Science bidang Pemasaran : Analisis Prilaku Pelanggan," *Data Sciences Indonesia (DSI)*, vol. 1, no. 1, hlm. 21–32, Nov 2021, doi: 10.47709/dsi.v1i1.1194.

[11] I. N. Rizki, D. Prayoga, M. L. Puspita, dan M. Q. Huda, "Implementasi Exploratory Data Analysis Untuk Analisis Dan Visualisasi Data Penderita Stroke Kalimantan Selatan Menggunakan Platform Tableau," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, Jan 2024, doi: 10.23960/jitet.v12i1.3856.

[12] I. Tiar, R. Muarif, dan D. Danar, "Implementasi Tableau Untuk Pengembangan VIsualisasi Data Pada Aplikasi Open Data Di Diskominfo Kabupaten Cirebon," 2024.

[13] A. M. Husein dan M. Harahap, "Pendekatan Data Science untuk Menemukan Churn Pelanggan pada Sector Perbankan dengan Machine Learning," *Data Sciences Indonesia (DSI)*, vol. 1, no. 1, hlm. 8–13, Nov 2021, doi: 10.47709/dsi.v1i1.1169.

[14] A. Deli *dkk.*, "Analisis Segmentasi Anggaran Pemasaran dan Penjualan Produk di Industri Retail Menggunakan K-Means Clustering Berbasis R Shiny," 2025.

[15] A. K. Ermy Pily, Susanti, U. Rio, dan Tashid, "Komparasi K-Means Clustering dengan Euclidean dan Cosine Similarity untuk Segmentasi dan Rekomendasi Produk pada Data E-Commerce," *The Indonesian Journal of Computer Science*, vol. 14, no. 2, Apr 2025, doi: 10.33022/ijcs.v14i2.4713.

[16] L. Azzahra dan A. Yasir, *Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food K-Means Clustering Method in Grouping Sales of Frozen Food Products*, vol. 2024, no. 1. 2024, hlm. 1–10. [Daring]. Tersedia pada: https://jurnal.unity-academy.sch.id/index.php/jirsi/index