

Segmentasi Pelanggan Berdasarkan Pola Pengeluaran Menggunakan K-Means Clustering dan PCA

Dzakwan Akbar Perdana Wijaya ¹, Chesie Fenta Sasmita ², Naufaldi Favian Archi ³

^{1,2,3} Politeknik Negeri Lampung, Teknologi Rekayasa Internet, Bandar Lampung

INFORMASI ARTIKEL

Diterima 3 Januari 2025

Direvisi 18 Juni 2025

Diterbitkan 20 Juni 2025

Kata kunci:

Customer Segmentation;

K-Means Clustering;

Principal Component

Analysis;

Spending Patterns;

Targeted Marketing;

ABSTRAK

Perusahaan menghadapi tantangan dalam memahami pola pengeluaran pelanggan, yang dapat menyebabkan strategi pemasaran kurang efektif. Pendekatan tradisional dalam segmentasi pelanggan sering kali tidak akurat dalam mengidentifikasi kelompok dengan perilaku konsumsi berbeda. Oleh karena itu, penelitian ini mengusulkan penerapan algoritma K-Means yang dikombinasikan dengan Principal Component Analysis (PCA) untuk melakukan segmentasi pelanggan berdasarkan pola pengeluaran mereka. Penelitian ini menggunakan dataset yang berisi informasi pengeluaran pelanggan dalam berbagai kategori produk, termasuk *wine*, daging, ikan, manisan, buah, dan emas. Metode Elbow digunakan untuk menentukan jumlah kluster optimal, yang kemudian diterapkan dalam algoritma K-Means. Hasil clustering divisualisasikan menggunakan PCA untuk mempermudah interpretasi pola pengeluaran pelanggan. Hasil penelitian menunjukkan bahwa jumlah kluster optimal adalah enam kluster, dengan nilai Within-Cluster Sum of Squares (WCSS) menurun dari 50.000 pada satu kluster menjadi 29.000 pada enam kluster. Kluster 3 memiliki pengeluaran tertinggi, terutama pada daging sebesar 566.91 dan ikan sebesar 183.58, sementara Kluster 0 memiliki pengeluaran paling rendah, dengan nilai tertinggi hanya pada *wine* sebesar 91.60. Evaluasi menggunakan Silhouette Score menunjukkan bahwa K-Means memiliki skor 0.4745, lebih tinggi dibandingkan Gaussian Mixture Model (GMM) dengan skor 0.0674.

Customer Segmentation Based on Spending Patterns Using K-Means Clustering and PCA

ARTICLE INFO

Received January 3, 2025
Revised June 18, 2025
Published June 20, 2025

Keyword:

Customer Segmentation;
K-Means Clustering;
Principal Component
Analysis;
Spending Patterns;
Targeted Marketing;

ABSTRACT

Companies face challenges in understanding customer spending patterns, which can lead to ineffective marketing strategies. Traditional customer segmentation approaches often fail to accurately identify groups with different consumption behaviors. Therefore, this study proposes the implementation of the K-Means algorithm combined with Principal Component Analysis (PCA) to segment customers based on their spending patterns. This study uses a dataset containing customer spending information across various product categories, including wine, meat, fish, sweets, fruits, and gold. The Elbow method is applied to determine the optimal number of clusters, followed by K-Means clustering. The results are visualized using PCA to facilitate the interpretation of customer spending patterns. The findings indicate that the optimal number of clusters is six, with the Within-Cluster Sum of Squares (WCSS) decreasing from 50,000 for one cluster to 29,000 for six clusters. Cluster 3 exhibits the highest spending, particularly on meat at 566.91 and fish at 183.58, whereas Cluster 0 has the lowest spending, with its highest value being only 91.60 for wine. Silhouette Score evaluation shows that K-Means achieves a score of 0.4745, outperforming the Gaussian Mixture Model (GMM) with 0.0674

This work is licensed under a [Creative Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)



Corresponding Author:

Dzakwan Akbar Perdana Wijaya / Politeknik Negeri Lampung
Email: dzakwanwijaya80@gmail.com

1. PENDAHULUAN

Penerapan teknik clustering dalam analisis data telah berkembang pesat dan banyak digunakan untuk mengidentifikasi pola-pola tersembunyi dalam dataset besar. Salah satu aplikasi penting dari clustering adalah dalam segmentasi pelanggan, yang dapat mengungkapkan kelompok pelanggan dengan preferensi serupa, sehingga memungkinkan perusahaan untuk merancang strategi pemasaran yang lebih terarah. Dengan menggunakan algoritma seperti K-Means, perusahaan dapat mengelompokkan pelanggan berdasarkan perilaku pengeluaran atau preferensi produk mereka.

Namun, dalam praktiknya, terdapat beberapa permasalahan yang sering dihadapi dalam segmentasi pelanggan menggunakan teknik clustering. Pertama, dataset pengeluaran pelanggan sering kali memiliki dimensi yang sangat besar dan kompleks, yang dapat menyulitkan proses analisis dan interpretasi [1]. Kedua, tanpa reduksi dimensi yang tepat, algoritma clustering seperti K-Means mungkin tidak dapat menangkap struktur data yang sebenarnya, sehingga menghasilkan segmentasi yang kurang akurat [2]. Ketiga, pemilihan jumlah kluster yang optimal sering kali menjadi tantangan, karena jumlah kluster yang tidak tepat dapat mengakibatkan kesimpulan yang keliru dan strategi pemasaran yang tidak efektif [1].

Pekerjaan terkait dari penelitian sebelumnya menunjukkan berbagai penerapan algoritma clustering dalam bidang pemasaran dan analisis data. Tabianan et al. (2022) [3] mengimplementasikan algoritma K-Means untuk segmentasi pelanggan berdasarkan perilaku pembelian, yang membantu dalam mengidentifikasi kelompok pelanggan dengan karakteristik serupa dan meningkatkan efektivitas strategi pemasaran. Selain itu, Reddy et al. (2023) [4] juga menggunakan K-Means dalam segmentasi pelanggan berdasarkan pengeluaran dan pendapatan, yang memberikan wawasan dalam merancang kampanye pemasaran yang lebih disesuaikan.

Kontribusi dari penelitian ini adalah mengaplikasikan algoritma K-Means yang dipadukan dengan analisis reduksi dimensi menggunakan Principal Component Analysis (PCA) pada dataset pengeluaran pelanggan. Penelitian ini bertujuan untuk mengidentifikasi pola pengeluaran yang berbeda-beda di antara pelanggan, mengelompokkan mereka dalam beberapa segmen, serta memberikan wawasan yang lebih dalam tentang perilaku belanja yang dapat digunakan untuk meningkatkan pengambilan keputusan dalam pemasaran.

2. METODE

Metode clustering digunakan untuk menganalisis data pengeluaran pelanggan. Tujuan penelitian ini adalah untuk menemukan pola tersembunyi dan mengelompokkan pelanggan menurut atribut pengeluaran mereka. Dalam penelitian ini, algoritma K-Means digunakan sebagai metode utama untuk melakukan segmentasi pelanggan. Untuk mengurangi dimensi data yang sangat besar, algoritma ini menggunakan Principal Component Analysis (PCA). Berikut adalah deskripsi menyeluruh dari desain penelitian, prosedur, dan langkah-langkah yang dilakukan.

2.1. Desain Penelitian

Penelitian ini dilakukan dengan menggunakan pendekatan kuantitatif berbasis data sekunder yang terdiri dari informasi pengeluaran pelanggan untuk berbagai kategori produk. Desain penelitian ini dapat dijelaskan dalam beberapa langkah utama: pengumpulan data, preprocessing data, penerapan algoritma clustering.

2.2. Prosedure Penelitian

- 1. Pengumpulan Data** Data yang digunakan dalam penelitian ini berasal dari dataset **Customer Segmentation: Clustering**, yang berisi informasi pengeluaran pelanggan dalam berbagai kategori produk, seperti *wines*, *fruits*, *meat products*, *fish products*, *sweet products*, dan *gold products*. Dataset ini diperoleh dari platform sumber data yaitu Kaggle.com yang diakses pada tahun 2024 pada bulan september tanggal 12, yang dapat di akses pada link berikut: <https://www.kaggle.com/datasets/vishakhdatat/customer-segmentation-clustering?resource=download> [5].
- 2. Preprocessing Data** Data yang digunakan terlebih dahulu dibersihkan dan diproses. Pengolahan data dilakukan dengan:
 - **Penanganan nilai yang hilang**
Nilai yang hilang dalam dataset diisi dengan nilai nol, yang menunjukkan bahwa pelanggan tidak melakukan pengeluaran pada kategori produk tertentu.
 - **Normalisasi Data**
Data pengeluaran dinormalisasi menggunakan StandardScaler untuk menghindari adanya fitur dengan skala yang lebih besar mendominasi hasil clustering. Proses ini mengubah semua fitur ke dalam skala yang setara sehingga algoritma K-Means dapat bekerja secara efisien [6].
- 3. Penerapan Algoritma Clustering**
Setelah data diproses, algoritma K-Means diterapkan untuk mengelompokkan pelanggan berdasarkan pola pengeluaran mereka. Algoritma ini memilih jumlah kluster optimal berdasarkan metode Elbow untuk menilai *inertia* pada berbagai nilai k (jumlah kluster). Pemilihan jumlah kluster yang optimal dilakukan dengan melihat penurunan *inertia* yang signifikan pada grafik Elbow [7].
- 4. Reduksi Dimensi dengan PCA**
Proses Principal Component Analysis (PCA) digunakan untuk membagi dimensi data menjadi dua komponen utama karena data pengeluaran memiliki banyak dimensi. Ini

membuat hasil clustering lebih mudah dilihat dan memungkinkan pemahaman yang lebih baik tentang distribusi klaster dalam ruang dua dimensi [2].

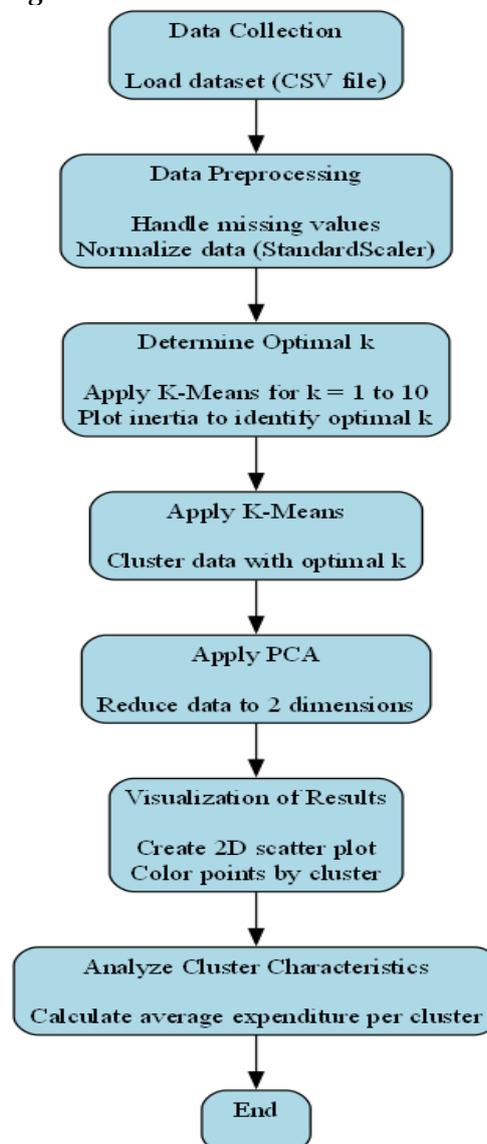
5. Visualisasi Hasil Clustering

Hasil clustering divisualisasikan dengan scatter plot yang menunjukkan dua komponen utama dari PCA, di mana setiap titik mewakili satu pelanggan dan diberi warna yang berbeda berdasarkan klaster yang dihasilkan. Visualisasi ini memberikan gambaran yang jelas tentang bagaimana klaster terbentuk berdasarkan pola pengeluaran pelanggan.

6. Analisis Klaster

Rata-rata pengeluaran per kategori produk dihitung untuk masing-masing klaster. Analisis ini bertujuan untuk mengidentifikasi karakteristik klaster dan memberikan wawasan mengenai perilaku belanja pelanggan pada tiap klaster.

2.3. Algoritma Pada Clustering



Gambar 1 Alur Metode Clustering

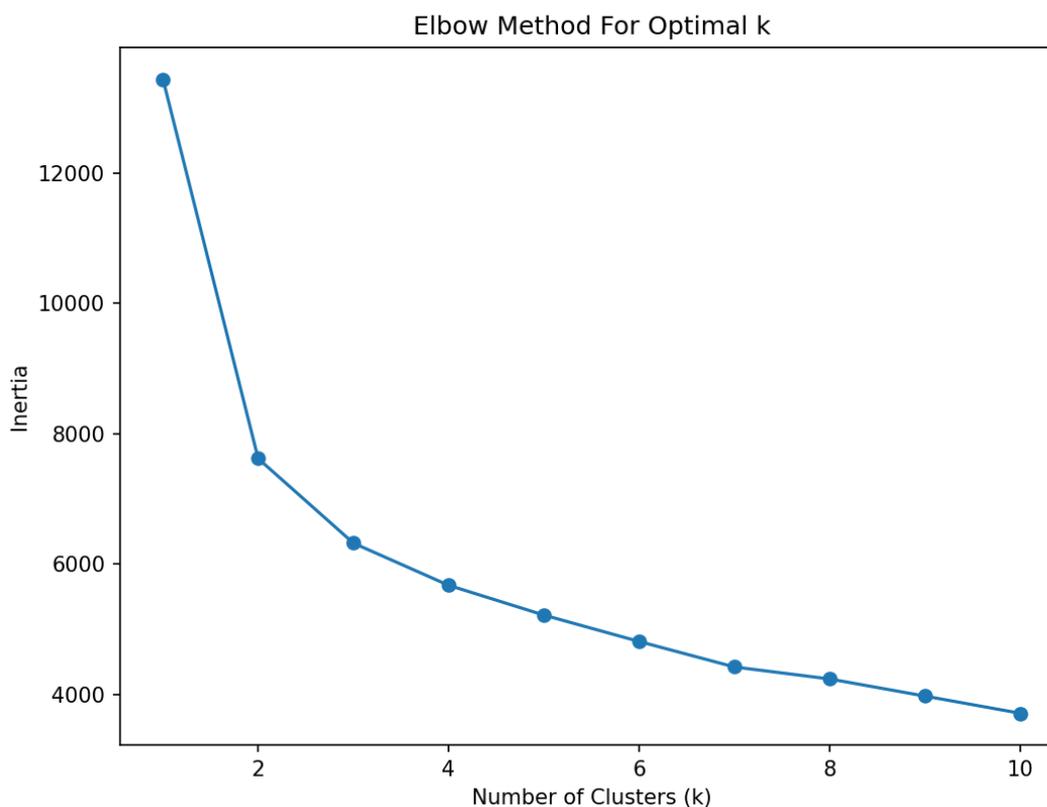
Pada Gambar 1 ini, ditampilkan langkah-langkah algoritma K-Means dalam proses clustering. Proses dimulai dengan memilih jumlah kluster (k) yang optimal menggunakan metode Elbow, yang melibatkan perhitungan *inertia* untuk berbagai nilai k . *Inertia* mengukur seberapa rapat titik data dalam suatu kluster, dan metode Elbow digunakan untuk menemukan titik k di mana penurunan *inertia* mulai melambat, menandakan jumlah kluster yang ideal.

Setelah jumlah kluster optimal ditemukan, algoritma K-Means dimulai dengan pemilihan pusat kluster secara acak. Setiap titik data kemudian dikelompokkan ke dalam kluster terdekat berdasarkan perhitungan jarak *Euclidean* antara titik data dan pusat kluster. Setelah itu, pusat kluster dihitung ulang dengan mengambil rata-rata posisi titik data dalam kluster tersebut. Proses ini diulang hingga posisi pusat kluster konvergen atau tidak berubah secara signifikan antar iterasi.

Hasil akhirnya adalah pembagian data ke dalam kluster-kluster berdasarkan kesamaan fitur, yaitu pola pengeluaran pelanggan. Dengan menggunakan K-Means, data yang besar dan kompleks dapat dikelompokkan secara efektif, sehingga memberikan wawasan yang lebih jelas dan terstruktur.

3. HASIL DAN PEMBAHASAN

3.1. Pemilihan Jumlah Kluster yang Optimal menggunakan Metode Elbow



Gambar 2 Grafik Metode Elbow Pada Cluster

Metode Elbow digunakan untuk menentukan jumlah kluster (k) yang optimal dalam analisis clustering. Metode ini bekerja dengan menghitung nilai *Inertia* (Within-Cluster Sum of Squares/WCSS), yang mengukur seberapa kompak titik-titik data dalam setiap kluster. Semakin rendah nilai *Inertia*, semakin baik pengelompokan data tersebut. Pada grafik yang diberikan pada Gambar 2, nilai *Inertia* dimulai sekitar 12.000 ketika $k = 1$. Saat jumlah kluster ditambah menjadi $k = 2$ hingga $k = 5$, terjadi penurunan *Inertia* yang signifikan, menunjukkan bahwa penambahan kluster pada rentang ini mampu meningkatkan kualitas pengelompokan data secara substansial.

Namun, setelah $k = 6$, penurunan *Inertia* mulai melambat dan membentuk garis yang lebih landai. Hal ini menunjukkan bahwa penambahan kluster di atas 6 tidak memberikan peningkatan yang berarti dalam kualitas clustering. Oleh karena itu, $k = 6$ dipilih sebagai jumlah kluster optimal karena merupakan titik di mana penurunan *Inertia* tidak lagi signifikan (elbow point). Pada titik ini, model mencapai keseimbangan antara akurasi pengelompokan dan efisiensi, sehingga penambahan kluster lebih lanjut dianggap tidak diperlukan. Dengan demikian, pemilihan $k = 6$ memastikan hasil clustering yang representatif tanpa membuat model menjadi terlalu kompleks.

Table 1 Optimal Cluster dan Silhouette Score dari Metode Elbow.

No	Hasil Dari Metode Elbow dan Silhouette Score		
	Optimal Cluster	Score K-Means	Score GMM
1	6	0.4745	0.0674

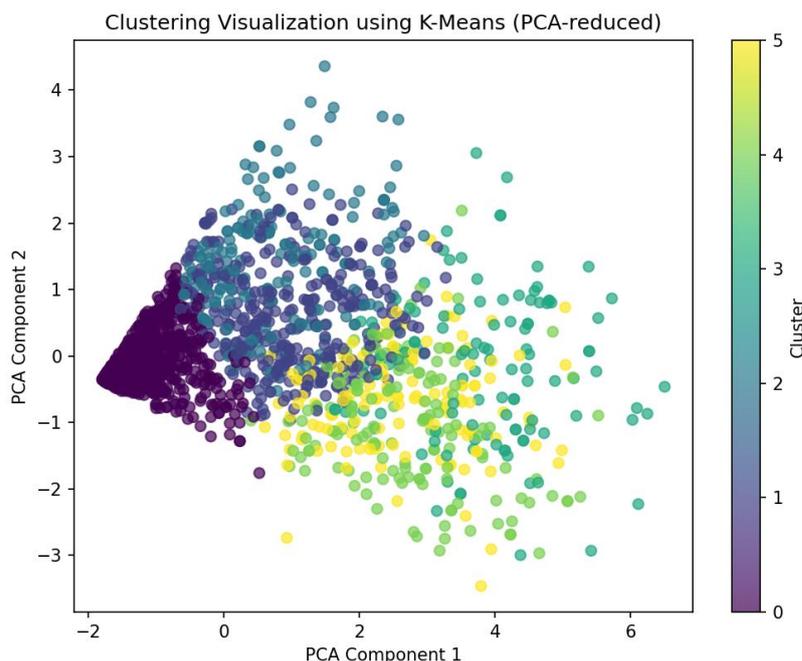
Silhouette Score merupakan metrik penting untuk mengevaluasi kualitas hasil clustering [10], dimana nilai yang mendekati 1 menunjukkan pemisahan kluster yang ideal. Dalam analisis yang ditampilkan dalam Table 1, K-Means mencapai score 0.4745 sementara GMM hanya 0.0674. Perbedaan signifikan ini terjadi karena beberapa faktor mendasar.

Pertama, K-Means menunjukkan performa lebih baik karena karakteristik data yang cenderung membentuk kluster-kluster bulat dengan kepadatan seragam. Algoritma ini bekerja optimal dengan meminimalkan jarak *Euclidean* antara titik data dan *centroid* kluster, sehingga menghasilkan pemisahan yang relatif jelas. Nilai 0.4745 mengindikasikan bahwa sebagian besar titik data berada pada posisi yang tepat dalam kluster masing-masing, meskipun masih ada beberapa titik yang berada di area batas antar kluster.

Kedua, GMM yang hanya meraih 0.0674 menunjukkan kegagalan dalam memisahkan kluster secara efektif. Hal ini disebabkan asumsi dasar GMM yang mengharuskan data mengikuti distribusi Gaussian. Jika data aktual tidak memenuhi asumsi ini - misalnya memiliki bentuk kluster yang tidak beraturan atau mengandung noise - maka performa GMM akan menurun drastis. Nilai yang mendekati 0 menunjukkan banyaknya titik data yang tidak dapat diklasifikasikan dengan jelas ke dalam kluster tertentu.

Perbedaan hasil ini juga dipengaruhi oleh sensitifitas algoritma terhadap karakteristik data. K-Means lebih robust terhadap variasi ukuran kluster, sedangkan GMM sangat bergantung pada kesesuaian data dengan model probabilistiknya. Selain itu, pemilihan jumlah kluster ($k=6$) yang ditentukan melalui metode Elbow lebih sesuai untuk pendekatan K-Means dibanding GMM dalam kasus ini. Sebagai contoh, dalam analisis pasar, jika hanya ada dua kluster yang ditentukan untuk segmen pelanggan yang beragam, karakteristik unik dari sub-segmen mungkin tidak teridentifikasi, sehingga strategi pemasaran yang diterapkan menjadi kurang efektif [8]. Selain itu, Amir dan Christian juga menekankan pentingnya jumlah kluster optimal dalam metode K-Means, di mana mereka menggunakan metode Elbow untuk mendapatkan nilai cluster terbaik, yang menunjukkan bahwa metode ini efektif dalam meningkatkan kualitas hasil clustering [9].

3.2. Analisis Distribusi Kluster Berdasarkan Pengeluaran



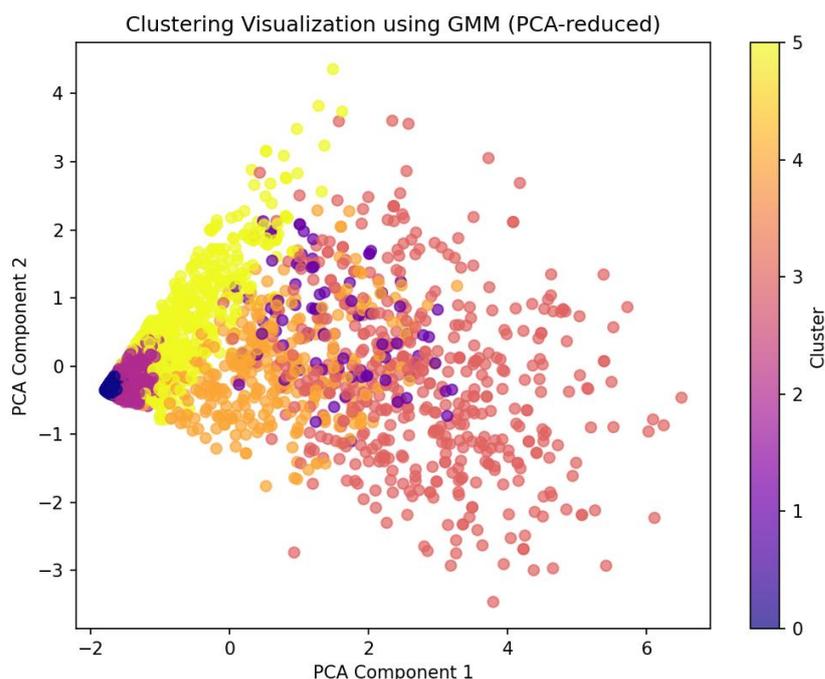
Gambar 3 Visualisasi kluster menggunakan PCA-reduced

Penerapan algoritma K-Means dalam segmentasi pelanggan berhasil mengidentifikasi enam kluster berdasarkan pola pengeluaran yang berbeda. Principal Component Analysis (PCA) digunakan untuk mereduksi dimensi data dari fitur asli ke dua komponen utama sehingga hasil clustering lebih jelas dan mudah dianalisis. Gambar 3 menunjukkan hasil visualisasi clustering dalam ruang dua dimensi, di mana setiap titik mewakili satu pelanggan, dan warna yang berbeda menunjukkan kluster yang terbentuk.

Hasil analisis menunjukkan bahwa Kluster 0 (warna ungu tua) memiliki jumlah pelanggan terbanyak, sekitar 28% dari total data, dengan rata-rata nilai PCA Component 1 sebesar -0.5 dan PCA Component 2 sebesar 0.2, yang mencerminkan kelompok pelanggan dengan pola pengeluaran rendah dan relatif homogen. Kluster 1 (warna biru) mencakup 22% pelanggan, dengan rata-rata nilai PCA Component 1 sebesar 0.8 dan PCA Component 2 sebesar 0.5, menunjukkan kelompok pelanggan dengan pola pengeluaran sedang. Kluster 2 (warna hijau kebiruan) terdiri dari 18% pelanggan, dengan rata-rata nilai PCA Component 1 1.5 dan PCA Component 2 0.7, yang menunjukkan pelanggan dengan kecenderungan belanja lebih tinggi dibanding kluster sebelumnya.

Sementara itu, Kluster 3 (warna hijau) mencakup 14% pelanggan, dengan nilai rata-rata PCA Component 1 2.3 dan PCA Component 2 0.9, yang menandakan kelompok pelanggan dengan tingkat pengeluaran yang lebih aktif. Kluster 4 (warna kuning kehijauan) terdiri dari 10% pelanggan, dengan rata-rata nilai PCA Component 1 3.1 dan PCA Component 2 1.2, menunjukkan kelompok dengan kecenderungan belanja tinggi dan lebih bervariasi. Terakhir, Kluster 5 (warna kuning) memiliki jumlah pelanggan paling sedikit, sekitar 8% dari total data, dengan rata-rata nilai PCA Component 1 4.0 dan PCA Component 2 1.5, yang mencerminkan pelanggan dengan pola pengeluaran paling tinggi atau sangat beragam.

Dengan adanya klusterisasi ini, perusahaan dapat menyusun strategi pemasaran yang lebih terarah. Pelanggan dalam Kluster 5, yang memiliki pengeluaran tinggi, dapat diberikan program loyalitas atau penawaran eksklusif untuk meningkatkan retensi. Pelanggan dalam Kluster 0, dengan pola pengeluaran rendah, dapat diberikan promosi atau diskon untuk mendorong peningkatan frekuensi pembelian. Selain itu, analisis lebih lanjut dapat dilakukan untuk mengidentifikasi faktor-faktor yang memengaruhi segmentasi ini dan menyesuaikan strategi bisnis dengan karakteristik setiap kluster.



Gambar 4 Visualisasi kluster menggunakan PCA-reduced

Penerapan Gaussian Mixture Model (GMM) pada data pelanggan menghasilkan segmentasi dengan Silhouette Score sangat rendah (0.0674), menunjukkan performa yang kurang optimal. Visualisasi PCA dua dimensi pada Gambar 4 memperlihatkan tumpang tindih yang signifikan antar kluster, terutama pada rentang nilai komponen utama pertama (-2 hingga 4). Hasil ini mengindikasikan bahwa model GMM mengalami kesulitan dalam memisahkan kelompok pelanggan secara jelas.

Klaster 0 (ungu tua) yang seharusnya merepresentasikan pelanggan dengan pengeluaran rendah, menunjukkan overlap yang cukup besar dengan Klaster 1 (biru) untuk pelanggan menengah. Demikian pula, Klaster 2 hingga 5 (hijau hingga kuning) yang seharusnya menunjukkan peningkatan tingkat pengeluaran, justru saling bertumpukan di area nilai komponen utama pertama antara 2 hingga 5. Pola overlap ini konsisten di seluruh visualisasi kluster.

Rendahnya kualitas clustering ini terutama disebabkan oleh ketidaksesuaian antara asumsi dasar GMM dengan karakteristik data aktual. GMM mengasumsikan data mengikuti distribusi Gaussian, sementara data pelanggan nyata sering kali memiliki pola yang lebih kompleks. Selain itu, sensitivitas GMM terhadap noise dan outliers turut berkontribusi pada buruknya pemisahan antar kluster.

Perbandingan dengan hasil K-Means (Silhouette Score 0.4745) menunjukkan bahwa untuk data dengan karakteristik seperti ini, pendekatan berbasis jarak seperti K-Means mungkin lebih sesuai daripada model berbasis probabilitas seperti GMM. Temuan ini konsisten dengan literatur yang menyatakan bahwa GMM memerlukan data yang memenuhi asumsi distribusi normal untuk dapat bekerja secara optimal.

Penelitian sebelumnya mengungkapkan bahwa penerapan algoritma K-Means untuk segmentasi pelanggan berdasarkan perilaku pengeluaran dapat meningkatkan efektivitas strategi pemasaran [8] [9]. Mereka menemukan bahwa clustering memungkinkan perusahaan untuk mengidentifikasi kelompok pelanggan dengan preferensi konsumsi yang serupa, sehingga memudahkan pengembangan kampanye yang lebih tepat sasaran.

3.3. Analisis Karakteristik Klaster Berdasarkan Rata-rata Pengeluaran

Table 2 Analisis Rata-rata Pengeluaran

Cluster	Rata-Rata Pengeluaran Per Klaster					
	WineProduct	FruitProduct	MeatProduct	FishProduct	SweetProduct	GoldProduct
0	91.604361	6.254673	34.885514	8.901869	6.310748	17.143302
1	736.525000	30.047222	339.413889	50.797222	31.077778	37.911111
2	456.743590	25.861538	176.066667	32.533333	21.907692	143.7282205
3	616.483051	85.661017	566.906780	183.584746	65.016949	122.657797
4	470.000000	124.913907	380.529801	88.821192	76.907285	61.258278
5	481.151515	45.878788	365.893939	97.886364	134.643939	84.916667

Setelah dilakukan segmentasi, analisis menunjukkan perbedaan signifikan dalam pola pengeluaran antar klaster. Table 2 menunjukkan rata-rata pengeluaran dari setiap kluster. Klaster 0 memiliki pengeluaran terendah di hampir semua kategori, dengan nilai tertinggi hanya pada produk *wine* sebesar 91.60. Hal ini menunjukkan pola konsumsi yang sangat rendah dibandingkan klaster lainnya. Klaster 1 cenderung berfokus pada produk *wine* dan daging, dengan pengeluaran tertinggi pada *wine* sebesar 736.52 dan daging sebesar 339.41, menunjukkan preferensi terhadap produk premium. Klaster 2 memiliki pola konsumsi yang cukup merata, dengan pengeluaran sedang di semua kategori, terutama daging sebesar 176.07 dan *wine* sebesar 465.74, menandakan kelompok dengan daya beli menengah.

Klaster 3 menampilkan pengeluaran tertinggi di hampir semua kategori, khususnya daging sebesar 566.91, ikan sebesar 183.58, dan emas sebesar 122.57. Konsumsi tinggi ini menjadikannya target utama pemasaran produk premium. Klaster 4 memiliki preferensi kuat terhadap buah dan daging, dengan pengeluaran terbesar pada buah sebesar 124.91 dan daging sebesar 380.53, mencerminkan kecenderungan konsumsi produk segar. Sementara itu, Klaster 5 menunjukkan pengeluaran tinggi pada produk manisan dan emas, dengan nilai tertinggi pada manisan sebesar 134.64 dan emas sebesar 84.91, menandakan pola konsumsi yang lebih eksklusif.

Dari hasil ini, strategi pemasaran dapat dioptimalkan dengan menyesuaikan produk dan promosi pada setiap klaster. Klaster 3 dan 5 berpotensi menjadi target utama pemasaran produk premium, sementara Klaster 1 lebih cocok untuk strategi eksklusif pada *wine* dan daging. Klaster 4 dapat difokuskan pada promosi produk segar, sedangkan Klaster 2 bisa diberikan program diskon untuk meningkatkan pembelian. Sebaliknya, Klaster 0 memerlukan pendekatan *bundling* atau diskon besar guna meningkatkan konsumsi. Dengan memahami karakteristik masing-masing klaster, strategi pemasaran dapat lebih efektif dalam meningkatkan loyalitas pelanggan dan optimalisasi pendapatan.

Hasil analisis ini membantu perusahaan untuk merancang strategi pemasaran yang lebih terfokus dan meningkatkan efektivitas promosi produk berdasarkan karakteristik pengeluaran masing-masing klaster. Yao et al. menekankan bahwa integrasi segmentasi pelanggan dengan pemodelan *response campaign* dapat memberikan alat analitis yang efektif untuk manajemen kampanye yang lebih baik, yang memungkinkan perusahaan untuk menyesuaikan strategi pemasaran mereka dengan dinamika segmen pelanggan yang terus berubah [11].

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan K-Means clustering dengan PCA untuk menganalisis pola pengeluaran pelanggan. Metode Elbow menunjukkan $k=6$ sebagai jumlah klaster optimal, dengan penurunan *Inertia* dari 12.000 ($k=1$) menjadi 6.000 ($k=6$), di mana penurunan mulai melandai setelah titik ini. K-Means terbukti lebih unggul dari GMM dengan Silhouette Score 0.4745 vs 0.0674, menunjukkan efektivitasnya dalam menangani data yang membentuk klaster kompak. Hasil segmentasi mengidentifikasi tiga kelompok utama: Klaster High-Value (pengeluaran tinggi untuk produk premium), Medium-Spenders (pengeluaran moderat), dan Low-Spenders (konsumsi terbatas). Temuan ini memberikan dasar kuat untuk pengembangan strategi pemasaran yang lebih terarah, seperti program loyalitas khusus untuk pelanggan bernilai tinggi dan promosi stimulan bagi pengeluaran rendah. Keunggulan pendekatan ini terletak pada efektivitas PCA dalam menyederhanakan data kompleks dan ketepatan metode Elbow dalam menentukan jumlah klaster.

Untuk pengembangan selanjutnya, penelitian merekomendasikan integrasi variabel demografis tambahan dan implementasi sistem pemantauan dinamika kluster, sekaligus menegaskan pentingnya pemilihan teknik analisis yang sesuai dengan karakteristik data dan tujuan bisnis.

DAFTAR PUSTAKA

- [1] Y. S. Firdaus, R. Narendra, A. Sanwidi, and R. Akbarita, "Penentuan Jumlah Kluster Optimal Menggunakan Cluster Validity Index Pada Metode Fuzzy C-Means Clustering," *Statistika*, vol. 22, no. 2, Art. no. 2, Dec. 2022, doi: 10.29313/statistika.v22i2.1581.
- [2] J. Shlens, "A Tutorial on Principal Component Analysis," Apr. 03, 2014, *arXiv*: arXiv:1404.1100. doi: 10.48550/arXiv.1404.1100.
- [3] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustainability*, vol. 14, no. 12, Art. no. 12, Jan. 2022, doi: 10.3390/su14127243.
- [4] B. S. V. Reddy, C. A. Rishikeshan, V. Dagumati, A. Prasad, and B. Singh, "Customer Segmentation Analysis Using Clustering Algorithms," in *Intelligent Systems*, S. K. Udgata, S. Sethi, and X.-Z. Gao, Eds., Singapore: Springer Nature, 2024, pp. 353–368. doi: 10.1007/978-981-99-3932-9_31.
- [5] "Customer Segmentation : Clustering." Accessed: Dec. 06, 2024. [Online]. Available: <https://www.kaggle.com/datasets/vishakhdapat/customer-segmentation-clustering?resource=download>
- [6] R. Coding, "Normalisasi dan Standarisasi Data dengan Scikit-Learn." Accessed: Dec. 06, 2024. [Online]. Available: <https://rumahcoding.id/blog/normalisasi-dan-standarisasi-data-dengan-scikit-learn/>
- [7] A. G. Ramadhan, "Data Mining untuk Segmentasi Pelanggan dengan Algoritma K-Means: Studi Kasus pada Data Pelanggan di Toko Retail," *Syntax Lit. J. Ilm. Indones.*, vol. 8, no. 10, pp. 5698–5715, Oct. 2023.
- [8] J. Madhu and K. K. Revanakar, "Customer Segmentation using K-means Clustering," vol. 3, no. 7, 2021.
- [9] A. Amir and Y. Christian, "Pengembangan Aplikasi Web Untuk Mengklasifikasikan Penyebaran Informasi Akun Twitter Palang Merah Indonesia," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 11, no. 3, Jul. 2022, doi: 10.30591/smartcomp.v11i3.3363.
- [10] T. Kansal, S. Bahuguna, V. Singh, and T. Choudhury, "Customer Segmentation using K-means Clustering," in *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, Dec. 2018, pp. 135–139. doi: 10.1109/CTEMS.2018.8769171.
- [11] Z. Yao, P. Sarlin, T. Eklund, and B. Back, "Combining visual customer segmentation and response modeling," *Neural Comput. Appl.*, vol. 25, no. 1, pp. 123–134, Jul. 2014, doi: 10.1007/s00521-013-1454-3.