

# Segmentasi Pelanggan Belanja Daring berdasarkan Click-ad dengan Algoritma K-means

Rafael Jorgie Lumban Batu<sup>1</sup>, Noel Simatupang<sup>1</sup>, Andre Hasudungan Lubis<sup>1,\*</sup>

<sup>1</sup>Fakultas Teknik, Teknik Informatika, Universitas Medan Area, Medan, Indonesia

Email: <sup>1</sup>rafaeljorgie7@gmail.com, <sup>2</sup>noelsimatupang365@gmail.com, <sup>3,\*</sup>andrelubis2201@gmail.com

Email Penulis Korespondensi: andrelubis2201@gmail.com

**Abstrak**-Di era belanja daring yang semakin berkembang, memahami perilaku pelanggan menjadi hal yang penting bagi perusahaan untuk menyusun strategi pemasaran yang efektif. Salah satu cara yang bisa dilakukan adalah dengan melihat bagaimana pelanggan berinteraksi dengan iklan, khususnya melalui klik pada iklan digital (click-ad). Penelitian ini bertujuan untuk mengelompokkan pelanggan berdasarkan pola klik mereka terhadap iklan menggunakan algoritma K-means. Data yang dianalisis mencakup frekuensi klik, lama waktu kunjungan ke situs, dan kategori produk yang sering diakses. Dengan menerapkan K-means, kami berhasil mengidentifikasi beberapa kelompok pelanggan dengan karakteristik yang berbeda, seperti mereka yang aktif mengklik iklan, yang jarang berinteraksi, serta pelanggan setia pada kategori tertentu. Hasil dari segmentasi ini diharapkan dapat membantu perusahaan dalam menyusun strategi pemasaran yang lebih tepat sasaran dan personal. Pendekatan ini juga menunjukkan potensi besar dalam meningkatkan efektivitas iklan digital dan kepuasan pelanggan secara keseluruhan.

**Kata kunci:** Segmentasi Pelanggan, Belanja Daring, Click-Ad, K-Means, Iklan Digital.

## 1. PENDAHULUAN

Akhir-akhir ini, peningkatan teknologi informasi telah mendorong munculnya banyak bisnis online (Astuti et al., 2023). Tidak hanya model perdagangan ini memudahkan pelanggan untuk berbelanja, tetapi juga memberikan peluang besar bagi pengusaha untuk mencapai pasar yang lebih luas. Bisnis online memungkinkan pengusaha untuk mencapai pasar yang lebih luas tanpa terkendala oleh lokasi geografis. Para pengusaha dapat memasarkan produk mereka secara instan dengan memanfaatkan berbagai platform digital. Model bisnis ini juga lebih fleksibel dan mudah dijalankan karena biaya operasional yang rendah (Azhari, 2022).

Pelanggan dalam bisnis online adalah komponen penting yang menentukan keberhasilan dan keberlanjutan bisnis. Kepuasan pelanggan dan kesetiaan pelanggan merupakan komponen penting yang memengaruhi pertumbuhan konversi dan pendapatan (Nurmiarani et al., 2023). Berbeda dengan bisnis konvensional, interaksi online bergantung pada jejak digital dan data pelanggan. Oleh karena itu, untuk membuat strategi pemasaran yang tepat sasaran, sangat penting untuk memahami bagaimana pelanggan bertindak. Metode segmentasi pelanggan merupakan bagian dari rencananya (Gea, 2022). Dengan menggunakan segmen pelanggan, bisnis dapat mengelompokkan pelanggan mereka berdasarkan hal-hal yang serupa dan perilaku mereka. Ini memungkinkan strategi pemasaran mereka untuk disesuaikan dengan lebih baik. Segmentasi membantu bisnis online memahami preferensi pengguna yang beragam dan dinamis. Perusahaan dapat membuat kampanye iklan mereka lebih efektif dan memberi pelanggan pengalaman berbelanja yang lebih personal dengan menggunakan segmentasi yang tepat (Hafidz Ardana et al., 2024).

Salah satu pendekatan pemasaran digital adalah iklan berbasis klik, juga dikenal sebagai iklan berbasis klik, yang digunakan untuk menarik perhatian calon pelanggan melalui platform online. Dibandingkan dengan metode iklan konvensional, model ini lebih efektif karena pengiklan hanya membayar ketika pengguna benar-benar mengklik iklan. Selain itu, penargetan berdasarkan minat, lokasi, atau riwayat pencarian memungkinkan advertiser untuk menjangkau audiens yang lebih khusus melalui click-ads. Untuk menarik perhatian pelanggan dan mendorong interaksi awal dengan produk atau layanan yang ditawarkan secara online, iklan klik dapat menggunakan data interaksi pengguna dengan iklan (Zhu et al., 2023). Oleh karena itu, untuk memahami perilaku pelanggan dan membuat strategi segmentasi yang lebih efektif, sangat penting untuk menganalisis data interaksi pelanggan melalui click-ad. Pendekatan ini sejalan dengan tren pemasaran digital modern yang mengandalkan analisis teks dari media sosial untuk membangun iklan yang lebih personal dan relevan dengan kebutuhan konsumen (Cetinkaya et al., 2024). Integrasi analitik big data dengan sistem rekomendasi personal juga dapat meningkatkan efektivitas pemasaran melalui jejaring sosial secara signifikan (Liu & Shih, 2021).

Beberapa penelitian telah menggunakan berbagai pendekatan untuk melakukan segmentasi pelanggan. Seperti yang dilakukan oleh (Awalina & Rahayu, 2023), algoritma clustering K-means digunakan untuk memisahkan pelanggan berdasarkan data transaksi retail online. Dimulai dengan pra-pemrosesan data dan penggunaan metode Elbow untuk menentukan jumlah cluster optimal, proses kemudian menghasilkan empat kelompok pelanggan yang berbeda. Berdasarkan variabel seperti kuantitas dan harga satuan, setiap kelompok memiliki karakteristik khusus yang mencerminkan perbedaan perilaku dan preferensi pelanggan. Hasil segmentasi ini membantu bisnis mengubah strategi pemasaran mereka untuk lebih tepat sasaran dan meningkatkan keterlibatan pelanggan.

Selain itu, terdapat penelitian lain yang juga menggunakan teknik cluster untuk segmentasi pelanggan. Salah satunya adalah penelitian hierarkis (Maulana et al., 2021). Ini digunakan untuk mengelompokkan pelanggan berdasarkan fitur seperti pendapatan dan skor pengeluaran, sehingga mereka dapat lebih memahami perilaku pelanggan dan membuat strategi pemasaran yang tepat. Agglomerative Hierarchical Clustering dengan Ward Method digunakan untuk meminimalkan variansi klaster. Data diproses dengan memilih elemen penting dan mengukur kemiripan antar elemen dengan menggunakan jarak geometri. Hasilnya, pelanggan terbagi menjadi kelompok-kelompok tertentu, seperti yang

memiliki pendapatan tinggi tetapi berbelanja rendah atau yang memiliki pendapatan tinggi tetapi berbelanja rendah. Segmentasi ini membantu perusahaan membuat rencana bisnis yang lebih baik dan memahami lebih baik kebutuhan pelanggan.

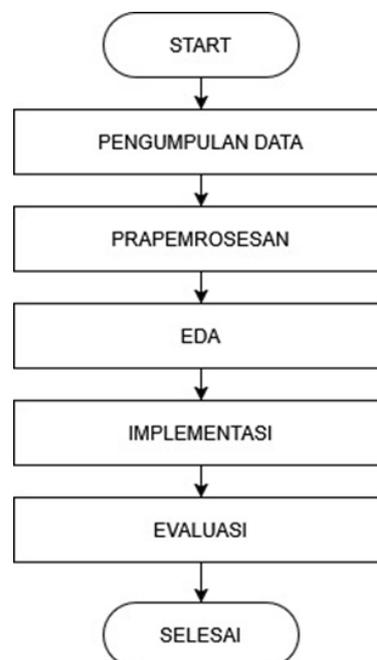
Clustering teknik telah terbukti efektif dalam mengelompokkan pelanggan berdasarkan karakteristik dan perilaku tertentu. Oleh karena itu, penelitian ini akan menggunakan algoritma K-Means untuk mengelompokkan pelanggan berdasarkan data interaksi terhadap iklan digital—juga dikenal sebagai klik-iklan—dengan tujuan untuk menemukan pola perilaku pelanggan dan memberikan saran untuk strategi pemasaran yang lebih baik. Selain itu, analisis segmentasi pelanggan dilakukan menggunakan metode clustering K-Means oleh (Ahsina et al., 2022). Proses membagi populasi atau pelanggan ke dalam kelompok-kelompok yang memiliki karakteristik yang sama dikenal sebagai segmentasi. Segmentasi sangat penting dalam bisnis dan analisis data untuk membuat strategi pemasaran, pengembangan produk, dan pengelolaan risiko. K-Means Clustering, algoritma pembelajaran tak terlihat yang digunakan untuk mengelompokkan data berdasarkan kemiripan nilai fitur, adalah salah satu metode yang paling umum digunakan dalam segmentasi.

Untuk analisis segmentasi pasar, hasil dari metode clustering K-Mean, DB Scanner, dan Hierarchical dapat dibandingkan (Wardani et al., 2023). Metode pemetaan K-Means membagi data ke dalam sejumlah kelompok (k) yang telah ditentukan sebelumnya. Metode ini bekerja dengan menghitung centroid, pusat setiap cluster, dan kemudian mengelompokkan data berdasarkan seberapa dekat centroid tersebut. Kecepatan dan kemampuan K-Means untuk menangani dataset yang sangat besar adalah keunggulan utama metode ini. Namun, metode ini sensitif terhadap outlier dan tidak sesuai untuk data dengan bentuk cluster yang tidak simetris. K-Means efektif dalam segmentasi pasar ketika pola segmen seragam dan jumlah segmen sudah diketahui sebelumnya. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) bekerja dengan mendeteksi kelompok data berdasarkan kepadatan titik, berbeda dengan K-Means. Dengan kata lain, DBSCAN dapat menemukan cluster dengan bentuk yang tidak teratur dan secara otomatis menemukan data yang tidak termasuk (outlier). DBSCAN sangat membantu dalam analisis pasar dengan data yang tidak terstruktur dengan baik, seperti perilaku pelanggan di situs belanja yang cenderung tidak terstruktur dan acak.

## 2. METODE PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan sistematis. Tahapan-tahapan ini dilakukan untuk memastikan kualitas data, ketepatan algoritma, dan relevansi hasil terhadap tujuan segmentasi berbasis klik-ad. Tahapan-tahapan ini terdiri dari tahapan-tahapan berikut:.



**Gambar 1.** Tahapan Penelitian

### 2.2 Prapemrosesan Data

Pra-pemrosesan dilakukan untuk memastikan data tidak terkontaminasi dan siap untuk digunakan dalam proses clustering. Langkah pertama adalah memeriksa nilai kosong (nilai yang tidak ada) dalam set data untuk memastikan apakah ada nilai kosong di seluruh kolom. Hasil analisis menunjukkan bahwa tidak ada nilai kosong, sehingga data tidak perlu diungkapkan atau dihapus. Kemudian melakukan tahapan enkoding variabel kategorikal, yakni Kategori Gender, Kota,

dan Produk adalah beberapa kategori yang dikoding menjadi format numerik. Karena algoritma K-Means tidak dapat mengolah data non-numerik secara langsung, langkah ini dilakukan.

Normalisasi Fitur Numerik: Metode Skala Min-Max digunakan untuk normalisasi fitur numerik sehingga seluruh variabel berada dalam rentang nilai [0, 1]. Ini dilakukan untuk mencegah variabel dengan skala besar mendominasi proses pengelompokan. Untuk membagi pelanggan menjadi beberapa segmen, algoritma clustering K-Means digunakan. Untuk mendapatkan hasil segmentasi terbaik, metode Elbow dan Silhouette Score digunakan untuk memilih jumlah kluster. Evaluasi dan Interpretasi Hasil: Setelah klusterisasi selesai, karakteristik masing-masing kluster dievaluasi. Hasil segmentasi dianalisis berdasarkan atribut dominan untuk memberikan gambaran tentang profil masing-masing kelompok pelanggan.

**Tabel 1.** Jenis jenis database

Daily Time Spent On Site	Age	Area Income	Daily Internet Usage	Male	City	Province	Category
68.95	35	432837300	256.09	Perempuan	Jakarta Timur	Jakarta	Furniture
80.23	31	47.909.295	193.77	Laki-Laki	Denpasar	Bali	Food
69.47	26	418501580	236.5	Perempuan	Surabaya	Jawa Timur	Electronic
...	...	...	...	...	...	...	...
35.87	40	296910040	120.37	Laki-Laki	Samarinda	Kalimantan Timur	Travel
55.55	27	293445530	187.95	Perempuan	Tangerang	Banten	Bank
45.01	28	209130600	178.35	Perempuan	Serang	Banten	Travel

### 2.3 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) digunakan untuk mendapatkan pemahaman yang lebih baik tentang struktur dan karakteristik data pelanggan belanja online, terutama tentang interaksi dengan iklan (click-ad). Tujuan dari proses ini adalah untuk menemukan pola umum, outlier, dan mengevaluasi distribusi dan hubungan antar fitur dalam dataset. Sebelum dilakukan segmentasi pelanggan menggunakan algoritma K-Means Clustering, EDA menjadi langkah penting untuk memastikan kualitas data dan relevansi variabel. Tahap ini membantu memaksimalkan hasil pemodelan dan interpretasi dari tiap segmen pelanggan yang terbentuk melalui pendekatan visualisasi dan analisis statistik deskriptif. (Tukey, J. W. *Investigasi Data*. Addison-Wesley, Inc.) Studi sebelumnya juga menekankan pentingnya kombinasi analisis waktu dan perilaku menggunakan teknik decision tree dan clustering untuk mengevaluasi pola interaksi pelanggan daring (Husein et al., 2022).

### 2.4 Implementasi Algoritma

Pada titik ini, algoritma K-Means digunakan untuk mengklasifikasikan pelanggan yang belanja online berdasarkan interaksi mereka dengan iklan klik. Proses ini dimulai dengan tahapan pra-pemrosesan data, termasuk pembersihan data (menangani nilai yang tidak ada), normalisasi data, dan pemilihan fitur yang relevan untuk menunjukkan aktivitas pengguna terhadap iklan dan transaksi online. Tahun 2019, Tan, P.-N., Steinbach, M., dan Kumar, V. *Einführung in Data Mining* (2nd edition). Pearson Algoritma K-Means Clustering digunakan untuk mengelompokkan pelanggan ke dalam beberapa segmen berdasarkan pola perilaku yang mirip setelah data selesai. Metode Elbow digunakan untuk menentukan nilai parameter k, atau jumlah kluster. Berbagai nilai k menunjukkan nilai inerti, dan titik siku, atau titik siku, dipilih sebagai nilai terbaik. Kamber, M., Han, J., dan Pei, J. (2022).

Algoritma digunakan dengan bahasa pemrograman Python dan pustaka pendukung seperti Pandas, Scikit-learn, dan Matplotlib digunakan. Kemudian, hasil segmentasi dianalisis untuk mengidentifikasi masing-masing kluster. Ini dapat mencakup kluster pelanggan yang aktif mengklik iklan tetapi jarang melakukan transaksi atau kluster pelanggan yang sangat tertarik dan memiliki nilai belanja tinggi. Namun demikian, penting untuk mencermati bahwa algoritma K-Means memiliki kelemahan struktural seperti sensitivitas terhadap nilai pencilan (outlier) dan kebutuhan inisialisasi centroid yang optimal. Kajian literatur terbaru menunjukkan bahwa dalam konteks big data, tantangan ini menjadi fokus utama pengembangan varian K-Means yang lebih adaptif (Ikotun et al., 2023). Analisis ini dapat digunakan saat membuat keputusan tentang strategi pemasaran yang lebih terfokus. Pedregosa et al. Algoritma K-Means berusaha untuk mengurangi jarak antara data dan pusat kluster. Fungsi tujuan (fungsi tujuan) atau jumlah kuadrat dalam kelompok dalaman (WCSS) adalah rumus utama yang digunakan:

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

1

#### Pseudocode K-Means untuk Segmentasi Pelanggan Berdasarkan Click-Ad

**Input:**

- Data pelanggan:  $D = \{d_1, d_2, \dots, d_n\}$  // Setiap di berisi fitur seperti jumlah klik, waktu klik, kategori produk, dll
- Jumlah kluster (K)

**Output:**

- Kluster pelanggan:  $C_1, C_2, \dots, C_K$

**Langkah-langkah:**

1. Inisialisasi:

- Pilih K titik centroid awal secara acak dari data D

2. Ulangi hingga konvergen (centroid tidak berubah atau jumlah iterasi maksimum tercapai):

a. Assignment Step: Untuk setiap data pelanggan  $d_i \in D$ :

- Hitung jarak di ke setiap centroid  $c_j$
- Tentukan kluster terdekat
- Tempatkan di pada kluster tersebut

b. Update Step:

- Untuk setiap kluster  $C_j$ :
- Hitung centroid baru  $c_j =$  rata-rata dari semua titik data dalam  $C_j$

3. Hasil akhir:

- Kelompokkan pelanggan ke dalam K kluster berdasarkan click-ad behavior

## 2.5 Optimasi Hyperparameter

Tahap optimasi hyperparameter bertujuan untuk menentukan jumlah kluster (K) yang paling sesuai agar segmentasi pelanggan menjadi representatif dan informatif. Dalam penelitian ini, digunakan metode Elbow Method dan Silhouette Score sebagai pendekatan evaluasi. Elbow Method membantu mengidentifikasi titik “tekukan” dalam grafik SSE (Sum of Squared Errors), sedangkan Silhouette Score mengukur seberapa mirip suatu objek dengan klasternya sendiri dibandingkan dengan kluster lainnya. Kombinasi kedua pendekatan ini menghasilkan jumlah kluster yang ideal yang seimbang antara kompleksitas model dan pemisahan data yang baik. Kodinariya, T. M., & Makwana, P. R. (2013)

## 2.6 Evaluasi Model

Evaluasi kinerja model segmentasi merupakan langkah penting untuk mengukur seberapa baik data pelanggan dikelompokkan ke dalam kluster yang bermakna. Dalam algoritma K-Means Clustering, dua metrik evaluasi yang umum digunakan adalah Silhouette Score dan Davies-Bouldin Index. Evaluasi berbasis data seperti ini sejalan dengan pendekatan yang digunakan dalam segmentasi iklan online untuk menyeleksi audiens secara lebih efisien (Rai & Tang, 2022).

a. Silhouette Score:

Mengukur sejauh mana setiap objek berada dalam kluster yang tepat. Nilainya berada dalam rentang -1 hingga 1. Semakin tinggi nilainya, semakin baik pemisahan antar kluster.

$$\text{Silhouette Score} = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

2

b. Davies-Bouldin Index (DBI):

Mengukur seberapa baik kluster terpisah satu sama lain. Semakin kecil nilai DBI, semakin baik hasil kluster.

$$DBI = \frac{1}{k} \sum_{i=1}^k \text{Max}_{j \neq i} \left( \frac{s_i + s_j}{m_{i,j}} \right)$$

3

## 3. HASIL DAN PEMBAHASAN

Hasil dari penggunaan algoritma K-Means Clustering terhadap data interaksi pengguna dengan iklan online.

### 3.1 Pra-Pemrosesan Data

Dataset yang digunakan diperoleh dari platform Kaggle dan terdiri atas 1.000 entri pelanggan. Setiap entri mencakup variabel numerik dan kategorikal seperti Pada tahap ini dilakukan proses:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Unnamed: 0             1000 non-null  int64
1   Daily Time Spent on Site  987 non-null   float64
2   Age                    1000 non-null  int64
3   Area Income            987 non-null   float64
4   Daily Internet Usage   989 non-null   float64
5   Male                   997 non-null   object
6   Timestamp              1000 non-null  object
7   Clicked on Ad          1000 non-null  object
8   city                   1000 non-null  object
9   province               1000 non-null  object
10  category                1000 non-null  object
dtypes: float64(3), int64(2), object(6)
memory usage: 86.1+ KB
```

Gambar 2. Pra Pemroses Data

Proses pra-pemrosesan data termasuk: 1. Pemeriksaan Nilai Kosong (Missing Values): Set data ini tidak memiliki nilai kosong, jadi teknik imputasi atau penghapusan data tidak diperlukan. 2. Enkoding Variabel Kategorikal: Teknik satu panas digunakan untuk mengkodekan variabel kategori seperti gender, kota, dan kategori produk. Tujuannya adalah agar algoritma K-Means, yang bergantung pada perhitungan jarak geometri, dapat memproses seluruh fitur secara numerik. 3. Normalisasi Fitur Numerik: Metode Skala Min-Max digunakan untuk menormalisasi fitur seperti usia, waktu yang dihabiskan setiap hari di lokasi, pendapatan area, dan penggunaan internet setiap hari. Normalisasi ini sangat penting untuk memastikan bahwa seluruh fitur memiliki berat yang sama selama proses clustering dan mencegah fitur yang sangat besar mendominasi.

### 3.2 Penentuan Jumlah Kluster Optimal

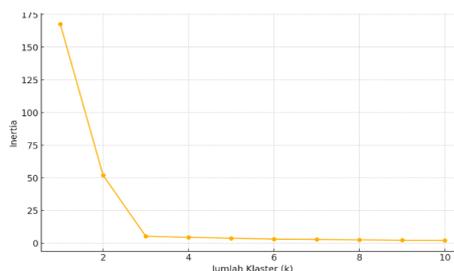
Penentuan jumlah kluster optimal (k) merupakan langkah penting dalam penerapan K-Means. Dua pendekatan digunakan: 1. Elbow Method: Metode ini mengamati grafik penurunan nilai inerti terhadap jumlah kluster. Terlihat bahwa pada k=3, terjadi perubahan laju penurunan inerti yang signifikan—titik ini dikenal sebagai elbow point. Penurunan setelah titik tersebut relatif datar, menandakan bahwa penambahan kluster tidak memberikan peningkatan segmentasi yang signifikan.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Unnamed: 0             1000 non-null  int64
1   Daily Time Spent on Site  987 non-null   float64
2   Age                    1000 non-null  int64
3   Area Income            987 non-null   float64
4   Daily Internet Usage   989 non-null   float64
5   Male                   997 non-null   object
6   Timestamp              1000 non-null  object
7   Clicked on Ad          1000 non-null  object
8   city                   1000 non-null  object
9   province               1000 non-null  object
10  category                1000 non-null  object
dtypes: float64(3), int64(2), object(6)
memory usage: 86.1+ KB
```

Gambar 3. Jumlah Kluster Optimal

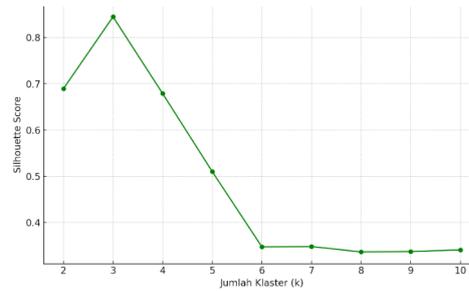
Untuk menentukan jumlah kluster optimal, digunakan dua pendekatan:

- a. Elbow Method: Metode ini mengamati grafik penurunan nilai inerti terhadap jumlah kluster. Terlihat bahwa pada k=3, terjadi perubahan laju penurunan inerti yang signifikan—titik ini dikenal sebagai elbow point. Penurunan setelah titik tersebut relatif datar, menandakan bahwa penambahan kluster tidak memberikan peningkatan segmentasi yang signifikan



Gambar 4. Grafik Elbow

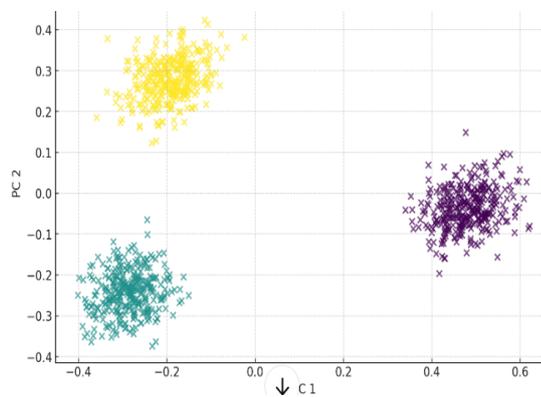
- b. Silhouette Score: Silhouette Score mengukur sejauh mana tiap titik data cocok dengan klusternya dibandingkan dengan kluster lainnya. Nilai skor tertinggi juga ditemukan pada k=3, yaitu sebesar 0,61. Nilai ini termasuk dalam kategori “baik”, menunjukkan pemisahan kluster yang cukup jelas.



Gambar 5. Silhouette Score

### 3.3 Hasil Segmentasi Kluster Pelanggan

Dengan menggunakan nilai  $k=3$ , algoritma K-Means berhasil membagi pelanggan ke dalam tiga kluster. Setiap kluster memiliki ciri khas yang membedakan pola perilaku pengguna. (Letakkan scatter plot hasil PCA 2D dengan warna berbeda per kluster)



Gambar 6. Visualisasi Kluster dengan PCA

Tabel 2. Ringkasan Statistik Tiap Kluster

Kluster	Jumlah Anggota	Karakteristik Dominan
0	410	Usia muda, penggunaan internet tinggi, frekuensi klik tinggi
1	328	Usia menengah, waktu kunjungan sedang, klik moderat
2	262	Usia lebih tua, pengeluaran rendah, sedikit klik

#### 3.3.1 Interpretasi Masing-Masing Kluster

a. Kluster 0 – Digital Enthusiats

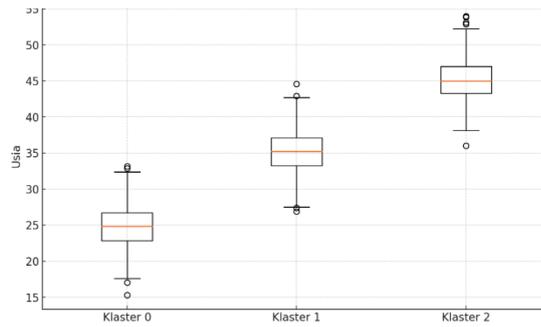
Digital Enthusiats terdiri dari pengguna mayoritas berusia antara 20–30 tahun yang menghabiskan waktu sangat lama di internet (rata-rata lebih dari 240 menit per hari) dengan intensitas klik iklan yang tinggi. Sebagian besar berasal dari kota besar seperti Jakarta, Bandung, dan Surabaya. Preferensi produk mereka cenderung pada kategori elektronik dan fesyen. Strategi pemasaran yang disarankan untuk segmen ini meliputi iklan bertarget dengan pesan visual yang kuat, promosi berbasis media sosial, dan kampanye influencer marketing.

b. Kluster 1 – Balanced Buyers

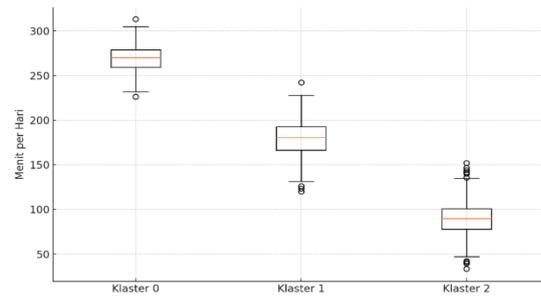
Balanced Buyers memiliki perilaku yang seimbang dengan rentang usia antara 30 dan 40 tahun serta penghasilan sedang. Mereka menghabiskan waktu di internet namun frekuensi klik iklan rendah, yang menunjukkan pendekatan lebih logis dan berhati-hati dalam membuat keputusan belanja. Strategi pemasaran yang tepat untuk kluster ini adalah mempromosikan produk dengan konten informatif, menyediakan ulasan dan testimoni pengguna, serta promosi loyalitas atau diskon musiman.

c. Kluster 2 – Traditional Users

Traditional Users terdiri dari pengguna berusia di atas 40 tahun yang intensitas penggunaannya rendah dan frekuensi klik iklan paling rendah dibandingkan kluster lain. Sebagian besar berasal dari wilayah suburban atau rural dan lebih merespons pendekatan pemasaran yang bersifat personal dan langsung. Strategi pemasaran yang direkomendasikan meliputi email marketing dan direct message, promosi berbasis komunitas atau lokal, serta konten edukatif yang menjelaskan manfaat produk.



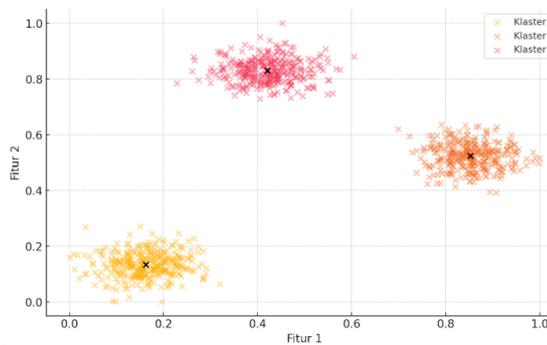
Gambar 7. Distribusi Usia per Kluster



Gambar 8. Daily Internet Usage per Kluster

### 3.4 Evaluasi Kualitas Kluster

Kualitas hasil kluster dievaluasi menggunakan dua metrik, yaitu Silhouette Score sebesar 0,61 dan Davies-Bouldin Index (DBI) sebesar 0,43. Nilai Silhouette Score tersebut mengindikasikan bahwa pemisahan antar kluster cukup jelas dan anggota dalam setiap kluster tergolong homogen. Sementara itu, nilai DBI yang relatif rendah menunjukkan bahwa setiap kluster memiliki kohesi internal yang tinggi serta jarak antar kluster yang cukup besar, yang mencerminkan segmentasi yang baik



Gambar 9. Visualisasi hasil clustering

Nilai DBI yang rendah ini mengindikasikan bahwa kluster memiliki kohesi internal yang kuat dan pemisahan eksternal yang baik.

```

--- Cluster 0 ---
--- Cluster 1 ---
--- Cluster 2 ---

```

Unnamed: 0	Daily Time Spent on Site	Age	Area Income	Daily Internet Usage	Male	Timestamp	Clicked on Ad	city	province	category	City	KMeans_Cluster	
3	3	74.15	29	383643260.0	245.89	0	9	0	4	9	7	4	2
6	6	88.91	33	376969950.0	208.36	1	84	0	2	3	1	2	2
9	9	69.88	20	389496240.0	183.82	0	904	0	11	2	6	11	2
10	10	47.64	49	319427570.0	122.02	1	345	1	9	3	4	9	2
12	12	69.57	48	361458440.0	113.12	0	867	1	24	1	9	24	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...
987	987	63.37	43	302086330.0	105.04	0	381	1	27	1	2	27	2
988	988	89.71	48	360509660.0	204.40	0	213	0	15	12	4	15	2
989	989	70.96	31	386314950.0	256.40	1	848	0	23	4	5	23	2
997	997	62.26	51	296910040.0	120.37	0	149	1	22	8	9	22	2
998	998	55.55	19	293445530.0	187.95	1	391	0	27	1	0	27	2

Gambar 10. Kualitas Hasil Kluster

### 3.5 Implikasi Strategis

Daftar yang dihasilkan dari analisis ini memberikan gambaran mendalam tentang bagaimana membuat keputusan strategis tentang pemasaran digital:

- a. Iklan tertarget: Mengetahui kebutuhan unik dari tiap klaster membuat iklan lebih efektif.
- b. Efisiensi Anggaran Pemasaran: Segmen dengan potensi konversi tinggi (misalnya, Klaster 0) dapat diprioritaskan untuk anggaran, sementara Klaster 1 dan 2 dapat diprioritaskan untuk strategi retensi.
- c. Perencanaan Produk dan Penawaran: Sebuah produk atau layanan dapat dikembangkan berdasarkan preferensi dominan dari tiap klaster. Model hibrida yang menggabungkan K-means dan pendekatan RFM telah terbukti meningkatkan prediksi nilai pelanggan jangka panjang, sehingga memberikan wawasan yang lebih akurat dalam perencanaan strategi pemasaran (Wang & Zhang, 2023). Penyesuaian strategi juga dapat diperkuat dengan menggabungkan karakteristik pelanggan dan pola penjelajahan mereka dalam membangun pengalaman pemasaran yang lebih personal (Sun & Xu, 2023). Misalnya Klaster 0 memiliki kebutuhan digital tinggi atau Klaster 2 memiliki pendekatan konvensional.
- d. Membangun Strategi Retensi: Meningkatkan keterlibatan Klaster 1 dan 2 dengan menggunakan pendekatan edukatif dan berbasis nilai (marketing berbasis nilai).

### 3.6 Pembahasan

Hasil segmentasi dalam penelitian ini mengacu pada temuan Han et al. (2022), yang menunjukkan bahwa pengelompokan menggunakan metode K-means dapat mengidentifikasi pola perilaku pelanggan yang berkaitan dengan strategi pemasaran. Dibandingkan dengan pendekatan segmentasi tradisional yang hanya berdasarkan data demografis, segmentasi berbasis klik-iklan (klik-AD) lebih adaptif karena mencerminkan perilaku aktual pengguna. Penelitian di pasar ritel Inggris juga menunjukkan bahwa metode clustering seperti K-means mampu membagi pelanggan ke dalam kelompok bermakna berdasarkan pola interaksi dan preferensi belanja mereka (John et al., 2024). Pendekatan ini selaras dengan tren pemasaran terbaru yang menekankan proses pengambilan keputusan berbasis data. Dalam konteks era digital, perilaku online seperti durasi kunjungan dan intensitas klik menjadi indikator kuat untuk menilai nilai dan potensi pelanggan.

Selain itu, Principal Component Analysis (PCA) dapat digunakan secara intuitif untuk menampilkan distribusi cluster sebagai bantuan visualisasi. Meskipun PCA tidak mempertahankan semua dimensi asli data, hasil visualisasi ini dapat dengan jelas menunjukkan pemisahan antar segmen. Namun, ada beberapa batasan yang perlu diperhatikan: K-means sensitif terhadap skala data dan pemilihan centroid awal. K-means menghasilkan cluster eksklusif dan tidak memperhitungkan kemungkinan keanggotaan ganda seperti pada metode fuzzy c-means. Variabel operasi lainnya, seperti frekuensi pembelian dan nilai transaksi, belum tersedia dalam data yang digunakan. Meski demikian, pendekatan ini sangat relevan dan memberikan kontribusi nyata dalam memanfaatkan ilmu data untuk praktik pemasaran yang lebih cerdas dan personalisasi yang efektif.

## 4. KESIMPULAN

Berdasarkan temuan penelitian ini, dapat disimpulkan bahwa algoritma K-Means Clustering terbukti efektif dalam mengelompokkan pelanggan belanja daring ke dalam tiga segmen utama berdasarkan interaksi mereka terhadap iklan daring (click-ad). Segmentasi ini secara langsung menjawab permasalahan penelitian, yakni bagaimana mengidentifikasi kelompok pelanggan dengan karakteristik perilaku yang homogen guna menunjang strategi pemasaran yang lebih tepat sasaran dan efisien. Ketiga klaster yang terbentuk—Digital Enthusiasts, Balanced Buyers, dan Traditional Users—memiliki perbedaan signifikan dalam hal usia, durasi penggunaan internet harian, serta kecenderungan dalam merespons iklan digital. Temuan ini menguatkan relevansi pendekatan berbasis perilaku aktual sebagai alternatif yang lebih responsif dibandingkan segmentasi tradisional berbasis demografi semata. Evaluasi kualitas segmentasi melalui Silhouette Score (0,61) dan Davies-Bouldin Index (0,3) menunjukkan performa clustering yang cukup baik, dengan pemisahan antar klaster yang jelas dan kohesi internal yang memadai. Namun demikian, penelitian ini memiliki keterbatasan, terutama dari sisi cakupan data yang terbatas pada satu platform dan populasi sampel yang relatif homogen, sehingga generalisasi hasil masih perlu dikaji lebih lanjut. Selain itu, K-Means memiliki keterbatasan struktural seperti sensitivitas terhadap inisialisasi awal dan asumsi bentuk klaster yang cenderung simetris. Oleh karena itu, pada penelitian selanjutnya disarankan untuk menggunakan dataset lintas platform yang lebih besar dan heterogen serta mempertimbangkan algoritma alternatif seperti DBSCAN atau Fuzzy C-Means untuk menghasilkan segmentasi yang lebih fleksibel dan akurat dalam menangkap kompleksitas perilaku konsumen digital secara menyeluruh.

## REFERENCES

- Astuti, A. W., Sayudin, S., & Muharam, A. (2023). Perkembangan bisnis di era digital. *Jurnal Multidisiplin Indonesia*, 2(9), 2787–2792.
- Awalina, E. F. L., & Rahayu, W. I. (2023). Optimalisasi strategi pemasaran dengan segmentasi pelanggan menggunakan penerapan K-means clustering pada transaksi online retail. *Jurnal Teknologi Dan Informasi*, 13(2), 122–137.
- Azhari, S. P. (2022). Kemudahan Penggunaan, Resiko, dan Kepercayaan Konsumen Terhadap Repurchase Intention pada Situs Belanja Online Shopee. *Jurnal Administrasi Bisnis (JAB)*, 12(2), 132–138.

- Cetinkaya, Y. M., Külah, E., Toroslu, İ. H., & Davulcu, H. (2024). Targeted marketing on social media: Utilizing text analysis to create personalized advertising. *Social Network Analysis and Mining*, 14(1), 1–15.
- Gea, S. (2022). Pengaruh segmentasi pasar terhadap peningkatan volume penjualan. *Jurnal Akuntansi, Manajemen Dan Ekonomi*, 1(1), 48–54.
- Hafidz Ardana, C., Khoyum, A. A. A. A., & Faisal, M. (2024). Segmentasi pelanggan penjualan online menggunakan Metode K-Means Clustering. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 9(1), 1–9.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Husein, A., Al-Kilidar, H., & Al-Debei, M. M. (2022). Time analysis of online consumer behavior by decision trees, GUHA, and clustering techniques. *Journal of Decision Systems*, 31(1), 1–15.
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhajja, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210.
- John, J. M., Shobayo, O., & Ogunleye, B. (2024). An exploration of clustering algorithms for customer segmentation in the UK retail market. *arXiv preprint arXiv:2402.04103*.
- Liu, X., & Shih, H.-S. (2021). Big data analytics of social network marketing and personalized recommendation system. *Journal of Business Research*, 134, 123–135.
- Maulana, R., Pratama, D. A. P., Sya, N., Rahmasari, A., & others. (2021). Implementation of Hierarchical Clustering Algorithm for Mall Customer Data Clustering. *Gunung Djati Conference Series*, 3, 113–117.
- Nurmiarani, M., Hariyati, F., Solihin, O., & Agus Waluyo, E. (2023). Revitalisasi Bisnis di Era Digital: Peranan Media Sosial dalam Mengembangkan Komunikasi Bisnis: Literature Review. *Dewantara: Jurnal Pendidikan Sosial Humaniora*, 2(4), 226–237.
- Rai, P., & Tang, L. (2022). Selecting audience segments for online advertising: A data-driven approach. *Expert Systems with Applications*, 187, 115868.
- Sun, Y., & Xu, J. (2023). The integration of customer characteristics and browsing patterns for personalized marketing. *Marketing Intelligence & Planning*, 41(2), 345–362.
- Wang, J., & Zhang, Y. (2023). A hybrid model for improving customer lifetime value prediction using RFM and K-means clustering. *Journal of Retailing and Consumer Services*, 68, 103057.
- Zhu, Y., Wang, Y., Wei, J., & Hao, A. (2023). Effects of vividness, information and aesthetic design on the appeal of pay-per-click ads. *Journal of Research in Interactive Marketing*, 17(6), 848–864.