

# **Implementasi K-Means Clustering Pada Segmentasi Kampanye Marketing Berdasarkan Karakteristik Audiens**



**Disusun oleh :  
Kelompok 9**

Muhammad Yusuf	122140193
Aldi Sanjaya	122140150
Lois Novel E Gurning	122140098
Nasywa Talitha Heryanna	122140046
Reynaldi Cristian Simamora	122140116

**Intelegensi Buatan  
Kelas RD**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI DAN INDUSTRI  
INSTITUT TEKNOLOGI SUMATERA  
2024**

## DAFTAR ISI

<b>DAFTAR ISI.....</b>	<b>2</b>
<b>BAB I LATAR BELAKANG.....</b>	<b>3</b>
<b>BAB II DASAR TEORI.....</b>	<b>5</b>
2.1 Marketing Campaign.....	5
2.2 Artificial Intelligence.....	5
2.4 Dataset.....	6
2.5. K-Means Clustering.....	6
2.6 Data Preprocessing.....	7
2.6.1 Data Cleaning.....	7
2.6.2 One Hot Encoding.....	7
2.6.3 Principal Component Analysis.....	7
2.7 Davies-Bouldin Index.....	8
<b>BAB III STUDI KASUS DAN PEMBAGIAN TUGAS.....</b>	<b>9</b>
3.1 Studi Kasus.....	9
3.2 Pembagian Tugas.....	10
<b>BAB IV IMPLEMENTASI CODE.....</b>	<b>11</b>
4.1 Implementasi Code.....	11
4.1.1 Code Bagian 1 (Menganalisis Nilai Dalam Column).....	11
4.1.2 Code Bagian 2 (Implementasi Algoritma K-Means Clustering).....	15
<b>BAB V KESIMPULAN DAN SARAN.....</b>	<b>23</b>
<b>DAFTAR PUSTAKA.....</b>	<b>24</b>
<b>LAMPIRAN.....</b>	<b>25</b>

# BAB I

## LATAR BELAKANG

Kemajuan teknologi telah mengubah cara perusahaan melakukan pemasaran, tidak hanya bergantung pada kreativitas, tetapi juga pada analisis data yang mendalam menggunakan teknologi informasi. Salah satu teknologi yang mendukung analisis ini adalah intelegensi buatan (AI), yang mengembangkan sistem yang meniru kecerdasan manusia untuk menyelesaikan tugas-tugas tertentu. Dalam AI, algoritma seperti *K-Means Clustering* digunakan untuk mengelompokkan data berdasarkan kesamaan atribut untuk membantu dalam segmentasi audiens. Algoritma ini memainkan peran penting dalam meningkatkan efisiensi dan akurasi sistem AI dalam berbagai aplikasi.

*K-Means* adalah algoritma *unsupervised learning* yang digunakan untuk mengelompokkan data tanpa memerlukan label atau informasi target. Tujuan utamanya adalah untuk membagi data menjadi beberapa grup atau cluster berdasarkan kesamaan fitur dalam data, sehingga objek dalam satu cluster lebih mirip satu sama lain dibandingkan dengan objek di klaster lain. Proses ini dimulai dengan memilih jumlah klaster ( $k$ ) yang diinginkan, kemudian algoritma menginisialisasi posisi acak centroid cluster dan mengelompokkan data berdasarkan jarak terdekat ke *centroid* tersebut. Proses ini diulang sampai posisi *centroid* stabil. Kalkulasi utama dalam *K-Means* adalah menghitung jarak antara setiap data dengan *centroid* menggunakan rumus jarak *Euclidean*, dan memperbarui posisi *centroid* berdasarkan rata-rata titik dalam klaster [1] [2].

Terdapat penelitian terdahulu yang menggunakan *K-Means Clustering* dan berfokus pada pengelompokan pasien berdasarkan kesamaan penyakit yang tercermin melalui kode *ICD*. Dalam penelitian ini, algoritma *K-Means* digabungkan dengan kesamaan semantik untuk mengukur kedekatan antar kode *ICD*, sehingga pengelompokan tidak hanya berdasarkan jarak data tetapi juga makna yang lebih mendalam. Outputnya adalah klasterisasi pasien berdasarkan kondisi medis yang serupa. Evaluasi dilakukan dengan metode *silhouette* untuk mengukur sejauh mana data terpisah antar klaster dan kedekatannya dalam klaster yang sama. Hasil penelitian menunjukkan peningkatan akurasi klasterisasi hingga 91,78% dengan penggunaan kesamaan semantik, dibandingkan dengan 84,93% tanpa pendekatan ini [3].

Rumusan masalah dalam penelitian ini adalah bagaimana karakteristik audiens dapat dijadikan indikator untuk menentukan strategi kampanye pemasaran yang tepat serta bagaimana mengevaluasi hasil pengelompokkan untuk memastikan kualitas segmentasi yang dihasilkan. Tujuan utamanya adalah melakukan pengelompokkan atau segmentasi kampanye pemasaran berdasarkan pola kinerja dan karakteristik audiens menggunakan algoritma *K-Means Clustering* dan mengevaluasi kualitas *clustering* tersebut menggunakan metode *Davies-Bouldin Index (DBI)*. Dengan mengelompokkan kampanye ke dalam *cluster* yang homogen, pemasar dapat mengidentifikasi strategi yang paling efektif, memahami preferensi audiens di berbagai segmen, dan mengoptimalkan alokasi sumber daya untuk meningkatkan ROI. Evaluasi menggunakan

DBI bertujuan untuk memastikan bahwa *cluster* yang dihasilkan memiliki pemisahan yang baik antar *cluster* dan kepadatan yang tinggi dalam masing-masing *cluster*, sehingga hasil segmentasi dapat diandalkan untuk menghasilkan wawasan berbasis data yang membantu pengambilan keputusan secara lebih terarah dan strategis.

Algoritma *K-means Clustering* pada penelitian ini diawali dengan menentukan 5 *cluster* dan mengolah *dataset* berisi 200.000 baris yang mencakup atribut seperti *Campaign\_Type*, *Target\_Audience*, *Duration*, dan *Channel\_Used*. Frekuensi unik dihitung untuk setiap kolom, diikuti dengan konversi data kategorikal menjadi format numerik menggunakan *One-Hot Encoding*. Model *K-Means* dilatih untuk mengelompokkan data, dan hasilnya divisualisasikan menggunakan *PCA* dalam *scatter plot*. Analisis menunjukkan 5 *cluster* dengan karakteristik unik yang menghasilkan segmentasi. Segmentasi ini memberikan wawasan strategis tentang kombinasi optimal jenis kampanye, audiens, durasi, dan saluran pemasaran untuk meningkatkan efektivitas kampanye.

Metode evaluasi yang digunakan pada penelitian ini adalah *Davies Bouldin Index (DBI)*. *Davies-Bouldin Index (DBI)* adalah salah satu metrik yang digunakan untuk mengevaluasi hasil clustering dalam model unsupervised learning, seperti *K-Means*. DBI mengukur seberapa baik cluster yang dihasilkan oleh model, dengan mempertimbangkan kedekatan antar centroid cluster dan sebaran dalam setiap cluster. Nilai *DBI* yang lebih rendah menunjukkan hasil clustering yang lebih baik, karena berarti cluster tersebut lebih terpisah dan lebih padat [4].

Setelah penerapan algoritma *K-Means*, hasil clustering memberikan strategi pemasaran yang lebih jelas untuk kelompok-kelompok kampanye yang memiliki karakteristik serupa. Setiap kampanye dikelompokkan ke dalam cluster yang sesuai, yang memungkinkan pemasar untuk mengidentifikasi pola-pola spesifik berdasarkan atribut seperti jenis kampanye, audiens yang ditargetkan, dan saluran yang digunakan. Analisis hasil clustering ini dapat digunakan untuk merancang strategi pemasaran yang lebih terfokus dan efisien, sesuai dengan segmen-segmen yang terbentuk berdasarkan karakteristik data yang relevan.

## **BAB II**

### **DASAR TEORI**

#### **2.1 *Marketing Campaign***

*Marketing campaign* merupakan rangkaian aktivitas berstrategi yang dilakukan dalam jangka waktu tertentu dan memiliki karakteristik audiens yang menjadi target untuk mencapai suatu tujuan spesifik. Pada umumnya, strategi yang dilakukan untuk mempromosikan produk, jasa, atau merek yang dikampanyekan dengan tujuan menarik perhatian audiens dan mendorong tindakan tertentu dari target audiens. Selain itu, *marketing campaign* juga dilakukan untuk menciptakan nilai bagi pelanggan dan membangun hubungan jangka panjang. *Marketing campaign* melibatkan berbagai saluran komunikasi seperti media sosial, email, iklan digital, televisi, dan radio, yang digunakan untuk menyampaikan pesan yang relevan kepada audiens yang telah ditentukan.

Terdapat beberapa faktor yang mempengaruhi suksesnya kampanye, yaitu [2]:

1. Penentuan tujuan yang jelas dan terukur sangat penting untuk menentukan strategi yang tepat.
2. Memahami karakteristik target audiens kampanye, sehingga dapat menentukan jenis kampanye yang cocok dan dapat menciptakan pesan yang lebih relevan dan menarik.
3. Konten yang baik, menarik, dan mudah dimengerti oleh target audiens dapat menarik perhatian audiens dan membangun minat kepada produk, jasa, atau merek yang dikampanyekan.
4. Memberikan instruksi yang jelas kepada audiens mengenai langkah-langkah yang ditawarkan membantu mengarahkan mereka ke arah yang diinginkan.

Faktor-faktor diatas dapat digunakan sebagai pertimbangan sebelum mulai melakukan kampanye. Strategi menjadi hal yang sangat krusial terhadap berjalannya dan hasil akhir dari kampanye. Memahami target audiens harus benar-benar diperhatikan, karena audiens menjadi salah satu tolok ukur suatu kampanye berhasil dan sukses. Dengan memahami target audiens, dapat dengan mudah menyusun strategi kampanye seperti pemilihan jenis kampanye yang akan dilakukan, media yang cocok digunakan, dan isi konten kampanye yang sesuai dan relevan dengan preferensi target audiens.

#### **2.2 *Artificial Intelligence***

Kecerdasan buatan atau *Artificial Intelligence* adalah kemampuan mesin untuk meniru dan melakukan tugas yang biasanya membutuhkan kecerdasan manusia. Dalam beberapa tahun terakhir, perkembangan komputasi dan pemrosesan data telah memungkinkan pengembangan algoritma dan sistem kecerdasan buatan yang semakin kompleks. Teknologi seperti pembelajaran mesin dan pembelajaran mendalam telah memungkinkan AI untuk belajar dari data dan meningkatkan kinerjanya dari waktu ke waktu. Kecerdasan buatan atau AI ini memiliki dampak bagi kehidupan manusia [5].

## 2.4 Dataset

Dataset adalah sekumpulan data yang terorganisir dan digunakan dalam analisis atau pelatihan model dalam pembelajaran mesin. Dataset digunakan sebagai dasar untuk membangun model yang dapat mengidentifikasi pola atau membuat prediksi berdasarkan data tersebut. Dataset dapat berbentuk tabel, gambar, teks, atau format lainnya, tergantung pada jenis data yang digunakan. Proses pengolahan dan analisis dataset melibatkan langkah-langkah seperti pembersihan data, transformasi, dan ekstraksi fitur untuk mempersiapkan data agar model dapat belajar dengan efektif [6].

## 2.5. K-Means Clustering

K-Means adalah suatu metode penganalisaan data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya.

Berikut adalah langkah-langkah dari algoritma K-Means:

1. Menentukan banyak k-cluster yang ingin dibentuk.
2. Membangkitkan nilai random untuk pusat cluster awal (centroid) sebanyak k-cluster.
3. Menghitung jarak setiap data input terhadap masing masing centroid menggunakan rumus jarak Euclidean (Euclidean Distance) hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Berikut adalah persamaan Euclidean Distance:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2}$$

Dimana:

$x_i$  : adalah data

$\mu_j$  : adalah centroid pada cluster j

Data akan dikelompokkan pada centroid dengan jarak terkecil memperbarui nilai centroid dengan nilai centroid yang baru didapatkan dari rata-rata cluster centroid tersebut dengan rumus :

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{x_j \in s_j} x_j$$

Dimana:

$\mu_j(t+1)$  : centroid baru pada iterasi  $(t+1)$

$N_{sj}$  : data pada cluster  $S_j$

4. Melakukan perulangan dari langkah 2 hingga 4 hingga anggota tiap cluster tidak ada yang berubah.
5. Jika langkah 4 telah terpenuhi, maka nilai rata-rata pusat cluster ( $\mu_j$ ) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data [7].

## **2.6 Data Preprocessing**

Data preprocessing adalah proses mengubah data mentah ke dalam bentuk yang lebih mudah dipahami. Proses ini diperlukan untuk memperbaiki kesalahan pada data mentah yang seringkali tidak lengkap dan memiliki format yang tidak teratur.

### **2.6.1 Data Cleaning**

Teknik data cleaning adalah suatu teknik yang digunakan untuk menangani data yang tidak lengkap. Proses data cleaning ini juga mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. [3]. Data Cleaning juga melakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal

### **2.6.2 One Hot Encoding**

One Hot Encoding adalah teknik dalam pengolahan data yang digunakan untuk mengubah variabel kategorikal menjadi bentuk numerik. Dalam One Hot Encoding, setiap nilai kategori diubah menjadi vektor biner yang hanya memiliki nilai 0 atau 1. Teknik ini sering digunakan dalam machine learning untuk memproses data yang memiliki variabel kategorikal.

### **2.6.3 Principal Component Analysis**

Principal component analysis (PCA) adalah suatu teknik analisis yang digunakan dalam bidang statistika dan juga data science. Dengan memanfaatkan teknik analisis ini, Anda dapat meringkas informasi yang tercantum dalam tabel data besar hingga menjadi beberapa kumpulan indeks ringkasan yang lebih kecil. Jadi, tahap visualisasi dan analisis data pun menjadi lebih mudah untuk dilakukan. Penerapan PCA sangat luas, bahkan hampir seluruh industri menerapkan teknik analisis ini. Dari industri kimia, misalnya, principal component analysis adalah teknik yang tepat untuk menggambarkan sifat sampel senyawa atau reaksi kimia tertentu.

## 2.7 Davies-Bouldin Index

Davies-Bouldin Index (DBI) adalah metrik evaluasi internal yang digunakan untuk mengevaluasi kualitas Clustering. Metrik ini mengukur seberapa jauh setiap Cluster berada dari Cluster lainnya, dengan nilai yang lebih rendah menunjukkan Clustering yang lebih baik [4]. DBI mempertimbangkan rata-rata jarak antara setiap pasangan Centroid Cluster dan ukuran relatif dari dua Cluster yang berdekatan.

Langkah penyelesaian Davies-Bouldin Index melibatkan beberapa langkah dasar, di antaranya adalah:

1. Menghitung Jarak Antar Centroid: Pertama, jarak antara setiap pasangan Centroid Cluster ( $d_{ij}$ ) dihitung menggunakan metrik jarak tertentu, seperti jarak Euclidean atau jarak Manhattan.
2. Menghitung Similitude: Selanjutnya, similitude antara dua Cluster yang berdekatan ( $R_{ij}$ ) dihitung sebagai jumlah jarak antara Centroid Cluster  $i$  dan  $j$  ( $d_{ij}$ ) dengan rata-rata jarak intra-Cluster  $i$  dan  $j$  ( $avg_i$  dan  $avg_j$ ):

$$R_{ij} = \frac{avg_i + avg_j}{d_{ij}}$$

3. Menghitung Davies-Bouldin Index: Kemudian, untuk setiap Cluster  $i$ , nilai Davies-Bouldin Index ( $DB_i$ ) dihitung sebagai maksimum dari similitude antara Cluster  $i$  dan Cluster lain:

$$DB_i = \max_{i \neq j} R_{ij}$$

4. Menghitung Davies-Bouldin Index Keseluruhan: Akhirnya, Davies-Bouldin Index keseluruhan (DB) diperoleh dengan mengambil rata-rata dari semua nilai  $DB_i$

$$DB = \frac{1}{k} \sum_i^k DB_i$$

Di mana  $k$  adalah jumlah total Cluster dalam Clustering.

Davies-Bouldin Index memberikan nilai yang lebih rendah untuk Clustering yang lebih baik, dengan nilai ideal mendekati nol. Semakin besar nilai Davies-Bouldin Index, semakin buruk kualitas Cluster [8].



## BAB III

### STUDI KASUS DAN PEMBAGIAN TUGAS

#### 3.1 Studi Kasus

Studi kasus yang diangkat yaitu segmentasi kampanye berdasarkan karakteristik audiens untuk meningkatkan efektivitas dalam pemasaran dengan menggunakan algoritma K-Means Clustering. Dataset yang digunakan berisi berbagai fitur seputar audiens yang dapat menggambarkan perilaku/preferensi berbeda-beda. Hal tersebut dapat membantu dalam mengidentifikasi kelompok audiens yang memiliki karakteristik serupa. Dengan melakukan clustering, kita dapat mengelompokkan audiens ke dalam beberapa segmen berdasarkan kemiripan dalam atribut-atribut yang relevan. Hal ini memungkinkan perusahaan untuk merancang kampanye pemasaran yang lebih terfokus dan efektif.

Algoritma K-Means bekerja dengan membagi data audiens ke dalam “K” kelompok, yang masing-masing memiliki ciri khas atau karakteristik tertentu yang mencerminkan pola dalam dataset [9]. Dalam tahap awal, K-Means akan menginisiasi posisi acak dari “K” centroid. Kemudian setiap titik data akan dipetakan ke cluster terdekat berdasarkan jarak dari titik data ke centroid. Proses ini akan diulang, memperbarui posisi centroid berdasarkan rata-rata posisi titik data yang terhubung dengan cluster tersebut, hingga mencapai perubahan minimal antara iterasi. Adanya segmentasi ini akan memudahkan perusahaan untuk kampanye pemasaran produk. Misalkan dengan hasil segmentasi, perusahaan dapat menyimpulkan bahwa campaign apa dan darimana penyampiannya, berapa lama durasinya, dan kepada target audiens tertentu.

Pada studi kasus ini kami menggunakan dataset yang terdiri atas 200.000 data, dan memiliki kolom-kolom seperti Campaign\_ID, Company, Campaign\_Type, Target\_Audience, Duration, Channel\_Used, Conversion\_Rate, Acquisition\_Cost, ROI, Location, Language, Clicks, Impression, Engagement\_Score, Customer\_Segment, dan Date. Kemudian dataset akan diolah untuk menentukan kolom-kolom mana yang relevan dalam kasus segmentasi kampanye marketing. Berdasarkan hasil pengolahan, maka dapat ditentukan kolom-kolom yang relevan yaitu kolom Campaign\_Type, Target\_Audience, Duration, dan Channel\_Used. Berikut adalah sampel dari dataset yang kami gunakan.

Tabel 3.1 Sampel Dataset

Campaign_ID	Company	Campaign_Type	Target_Audience	Duration	Channel_Used	Conversion_Rate	Acquisition_Cost	ROI	Location	Language	Clicks	Impression	Engagement_Score	Customer_Segment	Date
1	Innovate Industries	Email	Men 18-24	30 days	Google Ads	0.04	\$16,174.00	62.9	Chicago	Spanish	506	1922	6	Health & Wellness	2021-01-01
2	NexGen	Email	Women	60 days	Google Ads	0.12	\$11,566.00	51	New York	German	11	7523	7	Fashionistas	2021-01

	System s		35-44	ys				6 1		n	6				-02
3	Alpha Innovati ons	Influe ncer	Men 25-34	30 da ys	YouT ube	0.07	\$10,20 0.00	7 . 1 8	Los Ang eles	Fre nch	5 8 4	769 8	1	Outdoor Adventu rers	202 1-01 -03
4	DataTe ch Solutio ns	Displa y	All Ages	60 da ys	YouT ube	0.11	\$12,72 4.00	5 . 5 5	Mia mi	Ma nda rin	2 1 7	182 0	7	Health & Wellnes s	202 1-01 -04
5	NexGe n System s	Email	Men 25-34	15 da ys	YouT ube	0.05	\$16,45 2.00	6 . 5 0	Los Ang eles	Ma nda rin	3 7 9	420 1	3	Health & Wellnes s	202 1-01 -05

### 3.2 Pembagian Tugas

Tabel 3.2 Pembagian Tugas

Nama	Job Desk
Muhammad Yusuf	Mencari Dataset, Set-up Laporan, Mengimplementasikan studi kasus ke dalam Code Python, Mengerjakan Laporan BAB IV
Aldi Sanjaya	Mencari Dasar Teori dan membuatnya dalam laporan BAB II
Nasywa Talitha Heryanna	Membantu menambahkan dan melengkapi laporan BAB II Dasar Teori mengenai <i>Marketing Campaign</i> , mengerjakan BAB V
Lois Novel E Gurning	Menyusun BAB III, revisi laporan, PJ Asistensi
Reynaldi Cristian Simamora	Mencari jurnal berkaitan, menyusun BAB I

## BAB IV

### IMPLEMENTASI CODE

#### 4.1 Implementasi Code

Kode dikembangkan menggunakan bahasa Python, berikut merupakan syntax dari implementasi program yang telah dibuat `TubesIBKEL9_K-Means-Clustering.ipynb` :

##### 4.1.1 Code Bagian 1 (Menganalisis Nilai Dalam Column)

```
import pandas as pd

df = pd.read_csv('marketing_campaign_dataset.csv')

# Daftar nama kolom yang disebutkan
columns = [
    "Campaign_ID", "Company", "Campaign_Type", "Target_Audience",
    "Duration",
    "Channel_Used", "Conversion_Rate", "Acquisition_Cost", "ROI",
    "Location",
    "Language", "Clicks", "Impressions", "Engagement_Score",
    "Customer_Segment", "Date"
]

# Menyimpan hasil dalam list untuk setiap kolom
result = []

# Mengambil nilai unik dan frekuensinya
for column in columns:
    value_counts = df[column].value_counts()
    temp_df = pd.DataFrame({
        'Value': value_counts.index,
        'Frequency': value_counts.values
    })
    result.append(temp_df)

# Menampilkan hasil dalam bentuk tabel
for i, column in enumerate(columns):
    print(f"Frekuensi nilai dalam kolom {column}:")
    display(result[i])
    print("-" * 50)
```

Gambar 4.1 Code Bagian 1

#### A. Inisialisasi dataset

```
import pandas as pd

df = pd.read_csv('marketing_campaign_dataset.csv')
```

Gambar 4.2 Inisialisasi dataset

Pada Gambar 4.2 Inisialisasi dataset digunakan untuk menginisialisasi dataset csv yang sudah disiapkan untuk disimpan ke variabel df yang nantinya akan digunakan di dalam pengecekan isi column dataset.

#### B. Menyesuaikan nama kolom dengan dataset

```
# Daftar nama kolom yang disebutkan
columns = [
    "Campaign_ID", "Company", "Campaign_Type", "Target_Audience", "Duration",
    "Channel_Used", "Conversion_Rate", "Acquisition_Cost", "ROI", "Location",
    "Language", "Clicks", "Impressions", "Engagement_Score", "Customer_Segment",
    "Date"
]

# Menyimpan hasil dalam list untuk setiap kolom
result = []
```

Gambar 4.3 Menyesuaikan nama kolom dengan dataset

Gambar 4.3 diperuntukan untuk menyesuaikan nama kolom yang ada pada dataset, dan disimpan di list kosong variabel result untuk setiap kolom.

#### C. Mengambil Value & Frequency dari Column

```
# Mengambil nilai unik dan frekuensinya
for column in columns:
    value_counts = df[column].value_counts()
    temp_df = pd.DataFrame({
        'Value': value_counts.index,
        'Frequency': value_counts.values
    })
    result.append(temp_df)
```

Gambar 4.4 Mengambil Value & Frequency dari Column

Gambar 4.4 berfungsi untuk loop setiap column yang ada dan diambil value serta frequency dari tiap nilai yang sama agar bisa digunakan untuk membantu menganalisis studi kasus apa yang tepat untuk tipe dataset tersebut.

#### D. Memvisualkan dalam bentuk tabel agar mudah dilihat

```
for i, column in enumerate(columns):  
    print(f"Frekuensi nilai dalam kolom {column}:")  
    display(result[i]) # Jika kamu menggunakan Jupyter atau IPython  
    print("-" * 50)
```

Gambar 4.5 Memvisualkan dalam bentuk tabel agar mudah dilihat

#### E. Column yang dipakai

Dari semua kolom yang ditampilkan dan setelah melihat semua isi datanya, kolom yang relevan dengan studi kasus yang dibahas saat ini adalah *Campaign\_Type*, *Target\_Audience*, *Duration*, *Channel\_Load*. Pemilihan kolom-kolom *Campaign\_Type*, *Target\_Audience*, *Duration*, dan *Channel\_Used* dalam studi kasus "Segmentasi Kampanye Marketing Berdasarkan Karakteristik Audiens" didasarkan pada alasan-alasan berikut:

##### 1. Campaign\_Type

Kolom ini menunjukkan jenis kampanye yang dilakukan, seperti "Influencer", "Search", "Display", "Email", dan "Social Media". Tipe kampanye ini sangat relevan dalam segmentasi marketing karena dapat mempengaruhi cara audiens merespons dan terlibat dengan kampanye tersebut. Misalnya, audiens yang tertarik pada influencer marketing mungkin berbeda karakteristiknya dibandingkan audiens yang lebih responsif terhadap kampanye melalui iklan pencarian (search ads). Dengan mengelompokkan data berdasarkan jenis kampanye, kita bisa memahami jenis kampanye mana yang lebih efektif untuk audiens tertentu.

##### 2. Target\_Audience

Kolom ini mengidentifikasi kelompok audiens yang menjadi sasaran kampanye, seperti "Men 18-24", "Women 25-34", atau "All Ages". Segmentasi audiens berdasarkan demografi ini adalah kunci untuk memahami siapa yang lebih cenderung terlibat dengan kampanye tertentu. Dengan memahami karakteristik audiens, kita dapat menyesuaikan jenis kampanye, durasi, dan saluran yang digunakan untuk mencapai kelompok audiens yang paling relevan. Segmentasi ini akan membantu dalam menyesuaikan strategi pemasaran yang lebih tepat sasaran.

##### 3. Duration

Kolom ini menunjukkan durasi kampanye, seperti "30 days", "45 days", atau "60 days". Durasi kampanye sering kali mempengaruhi tingkat keterlibatan audiens dan efektivitas kampanye itu sendiri. Durasi yang lebih panjang mungkin memberikan audiens lebih banyak waktu untuk terlibat dengan kampanye, sementara durasi yang lebih singkat mungkin lebih efektif dalam menarik perhatian audiens secara cepat. Dalam

segmentasi, durasi kampanye dapat digunakan untuk melihat apakah audiens tertentu lebih responsif terhadap kampanye dengan durasi tertentu.

#### 4. Channel\_Used

Kolom ini menggambarkan saluran yang digunakan dalam kampanye pemasaran, seperti "Email", "Google Ads", "YouTube", "Instagram", atau "Website". Saluran yang digunakan dalam kampanye memiliki dampak besar pada jenis audiens yang akan terlibat. Misalnya, audiens yang aktif di media sosial mungkin lebih terlibat dengan kampanye yang menggunakan platform seperti Instagram atau Facebook, sementara audiens yang lebih tua atau profesional mungkin lebih cenderung berinteraksi dengan kampanye melalui email atau website. Segmentasi berdasarkan saluran ini memungkinkan untuk mengidentifikasi saluran pemasaran mana yang lebih efektif untuk audiens tertentu.

#### F. Output Code Bagian 1

Kolom-kolom lain dalam dataset, seperti *Acquisition\_Cost*, *Conversion\_Rate*, dan *ROI*, meskipun relevan untuk mengevaluasi performa kampanye setelah segmentasi, lebih berfokus pada hasil dan kinerja daripada karakteristik kampanye itu sendiri atau audiens yang terlibat. Untuk segmentasi audiens, kita lebih membutuhkan informasi tentang jenis kampanye, audiens yang ditargetkan, durasi kampanye, dan saluran yang digunakan, yang bisa memberikan wawasan lebih dalam tentang bagaimana kampanye seharusnya disesuaikan dengan karakteristik audiens yang berbeda. Dengan memilih kolom-kolom ini, kita bisa mengelompokkan data berdasarkan karakteristik audiens dan strategi kampanye yang digunakan, yang akan membantu dalam memahami dan mengoptimalkan segmentasi audiens dalam konteks kampanye pemasaran.



The image shows a Jupyter Notebook interface with two code cells. The first cell displays the frequency of values for the 'Campaign\_Type' column, and the second cell displays the frequency of values for the 'Target\_Audience' column. Both tables include an index column (0-4), the category name, and the frequency count.

	Value	Frequency
0	Influencer	40169
1	Search	40157
2	Display	39987
3	Email	39870
4	Social Media	39817

	Value	Frequency
0	Men 18-24	40258
1	Men 25-34	40023
2	All Ages	40019
3	Women 25-34	40013
4	Women 35-44	39687

Gambar 4.6 Output Value & Frequency Kolom Campaign Type & Target\_Audience

Frekuensi nilai dalam kolom Duration:

	Value	Frequency
0	30 days	50255
1	45 days	50100
2	60 days	49866
3	15 days	49779

---

Frekuensi nilai dalam kolom Channel\_Used:

	Value	Frequency
0	Email	33599
1	Google Ads	33438
2	YouTube	33392
3	Instagram	33392
4	Website	33360
5	Facebook	32819

Gambar 4.7 Output Value & Frequency Kolom Duration & Channel\_Used

#### 4.1.2 Code Bagian 2 (Implementasi Algoritma K-Means Clustering)

```
# Importing required Libraries
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
from sklearn.metrics import davies_bouldin_score

# --- 1. Data Preprocessing ---
# Inisialisasi Dataset
df = pd.read_csv('marketing_campaign_dataset.csv')

# Menampilkan apakah terdapat missing values dalam dataset
print("Apakah dataset ada nilai yang null?")
print(f"{df.isnull().sum()}")

# Menghapus baris dengan missing values
df_cleaned = df.dropna()

# Menampilkan apakah ada nilai yang null setelah pembersihan
print("\nSetelah menghapus missing values:")
print(f"{df_cleaned.isnull().sum()}")

# Pilih kolom yang relevan untuk analisis clustering
columns_to_use = ['Campaign_Type', 'Target_Audience', 'Duration',
'Channel_Used']
```

```

df_selected = df_cleaned[columns_to_use].copy()

# Melakukan One-Hot Encoding untuk data kategorikal
df_encoded = pd.get_dummies(df_selected)

# Menggunakan PCA untuk mengurangi dimensi data menjadi 2 komponen utama agar
dapat divisualisasikan
pca = PCA(n_components=2)
reduced_data = pca.fit_transform(df_encoded)

# --- 2. Building the K-means Model ---
# Tentukan jumlah cluster
kmeans = KMeans(n_clusters=5, init='k-means++', random_state=42)

# Latih model K-means menggunakan data yang sudah di-encode
kmeans.fit(df_encoded)

# Tambahkan hasil cluster ke dataframe
df_selected['Cluster'] = kmeans.labels_

# --- 3. Evaluating the Clustering Model ---
# Evaluasi model menggunakan Davies-Bouldin Index
db_score = davies_bouldin_score(df_encoded, kmeans.labels_)
print(f'Evaluasi Davies-Bouldin Index: {db_score}')

# --- 4. Visualizing the Results ---
# Visualisasi hasil clustering
plt.figure(figsize=(8, 6))
plt.scatter(reduced_data[:, 0], reduced_data[:, 1], c=kmeans.labels_,
            cmap='viridis')
plt.title('Hasil Clustering Kampanye Berdasarkan Karakteristik Audiens')
plt.xlabel('Pengaruh Jenis Kampanye dan Audiens')
plt.ylabel('Pengaruh Durasi dan Saluran Kampanye')
plt.colorbar(label='Cluster')
plt.show()

# Rangkuman jumlah data per cluster
print("Jumlah data per cluster:")
cluster_counts = df_selected['Cluster'].value_counts().sort_index()
print(cluster_counts)

# Menampilkan ringkasan nilai terbanyak (modus) dari setiap cluster
print("\nRangkuman nilai terbanyak (modus) dari setiap cluster:")
for cluster_id in sorted(df_selected['Cluster'].unique()):

```



```

cluster_data = df_selected[df_selected['Cluster'] == cluster_id]
print(f"\nCluster {cluster_id}:")
print("Jumlah Data:", len(cluster_data))
print("Nilai Terbanyak (Modus) untuk Setiap Kolom:")
print(cluster_data.mode().iloc[0])

```

Gambar 4.8 Code Bagian 2

#### A. Import Libraries

```

# Importing required libraries
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
from sklearn.metrics import davies_bouldin_score

```

Gambar 4.9 Pustaka yang diperlukan

Pada Gambar 4.9 yaitu bagian pertama kode ini, beberapa pustaka (*library*) yang diperlukan untuk menjalankan algoritma *K-Means clustering* dan visualisasi data diimpor. Pustaka *pandas* digunakan untuk memanipulasi dan memproses dataset, memungkinkan kita untuk membaca data, memfilter kolom, dan melakukan operasi lainnya pada data. Algoritma *K-Means clustering* diimplementasikan menggunakan *KMeans* dari *sklearn.cluster*, yang berfungsi untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kesamaan fitur. Selain itu, *PCA* dari *sklearn.decomposition* digunakan untuk mereduksi dimensi data, mengubah data yang memiliki banyak fitur menjadi hanya dua komponen utama agar dapat divisualisasikan dalam dua dimensi. Untuk keperluan visualisasi, *matplotlib.pyplot* digunakan untuk menghasilkan grafik dan scatter plot, yang memvisualisasikan hasil clustering dalam bentuk yang lebih mudah dipahami. Terakhir, untuk mengevaluasi hasil clustering, *davies\_bouldin\_score* dari *sklearn.metrics* digunakan, yang menghitung *indeks Davies-Bouldin* untuk mengukur kualitas clustering dengan menilai seberapa terpisah dan terpusat cluster yang terbentuk.

#### B. Data Cleaning

```

# --- 1. Data Preprocessing ---
# Inisialisasi Dataset
df = pd.read_csv('marketing_campaign_dataset.csv')

# Menampilkan apakah terdapat missing values dalam dataset
print("Apakah dataset ada nilai yang null?")
print(f"{df.isnull().sum()}")

# Menghapus baris dengan missing values
df_cleaned = df.dropna()

```

```

# Menampilkan apakah ada nilai yang null setelah pembersihan
print("\nSetelah menghapus missing values:")
print(f"{df_cleaned.isnull().sum()}")

# Pilih kolom yang relevan untuk analisis clustering
columns_to_use = ['Campaign_Type', 'Target_Audience', 'Duration',
'Channel_Used']
df_selected = df_cleaned[columns_to_use].copy()

# Melakukan One-Hot Encoding untuk data kategorikal
df_encoded = pd.get_dummies(df_selected)

# Menggunakan PCA untuk mengurangi dimensi data menjadi 2 komponen utama agar
dapat divisualisasikan
pca = PCA(n_components=2)
reduced_data = pca.fit_transform(df_encoded)

```

Gambar 4.10 Code Data Cleaning

Pada Gambar 4.10, dataset dibaca menggunakan `pd.read_csv()` dan disimpan dalam variabel `df`. Proses pertama yang dilakukan adalah memeriksa apakah terdapat nilai yang hilang (`missing values`) dalam dataset dengan menggunakan `isnull().sum()`, yang menghitung jumlah nilai `null` pada setiap kolom. Selanjutnya, baris-baris yang mengandung nilai yang hilang dihapus menggunakan `dropna()`, dan hasilnya disimpan dalam `df_cleaned`. Setelah pembersihan data, pengecekan kembali dilakukan untuk memastikan tidak ada nilai yang hilang.

Kemudian, bagian data yang relevan untuk clustering dipilih, yaitu kolom `Campaign_Type`, `Target_Audience`, `Duration`, dan `Channel_Used`. Kolom-kolom ini dipilih karena mereka mengandung informasi yang berguna untuk analisis clustering. Setelah itu, `pd.get_dummies()` digunakan untuk melakukan One-Hot Encoding terhadap data kategorikal, sehingga setiap kategori pada kolom-kolom tersebut diubah menjadi kolom biner (0 atau 1), dan hasilnya disimpan dalam `df_encoded`.

PCA digunakan untuk mengurangi dimensi data yang telah di-encode menjadi dua komponen utama (2D), sehingga hasil clustering dapat divisualisasikan secara grafis.

### C. Build K-Means Model

```

# --- 2. Building the K-means Model ---
# Tentukan jumlah cluster
kmeans = KMeans(n_clusters=5, init='k-means++', random_state=42)

# Latih model K-means menggunakan data yang sudah di-encode

```

```
kmeans.fit(df_encoded)

# Tambahkan hasil cluster ke dataframe
df_selected['Cluster'] = kmeans.labels_
```

Gambar 4.11 Code *Build K-Means Model*

Gambar 4.11 adalah inti dari algoritma K-Means clustering. Dimulai dengan menentukan jumlah cluster yang diinginkan, yang dalam hal ini adalah 5 cluster (`n_clusters=5`). K-Means diinisialisasi menggunakan `KMeans()` dengan parameter `init='k-means++'` yang memungkinkan pemilihan centroid awal yang lebih baik dan mengurangi kemungkinan hasil clustering yang buruk. Proses clustering dimulai dengan `kmeans.fit(df_encoded)`, di mana model K-Means melatih dirinya pada data yang telah di-encode.

Setelah model selesai dilatih, hasil prediksi cluster (`kmeans.labels_`) ditambahkan ke dataframe `df_selected` dalam bentuk kolom baru yang disebut 'Cluster'. Kolom ini menunjukkan ke cluster mana masing-masing data (baris) dikelompokkan.

#### D. Evaluating the Clustering Model

```
# --- 3. Evaluating the Clustering Model ---
# Menampilkan jumlah data per cluster
cluster_counts = df_selected['Cluster'].value_counts()
print("Jumlah data per cluster:")
print(cluster_counts)

# Evaluasi model menggunakan Davies-Bouldin Index
db_score = davies_bouldin_score(df_encoded, kmeans.labels_)
print(f'Evaluasi Davies-Bouldin Index: {db_score}')
```

Gambar 4.12 Code *Evaluating the Clustering Model*

Gambar 4.12 adalah evaluasi hasil clustering. Pertama, jumlah data yang terdapat di setiap cluster ditampilkan menggunakan `value_counts()` pada kolom 'Cluster'. Ini memberikan gambaran tentang distribusi data dalam cluster yang berbeda.

Selanjutnya, evaluasi dilakukan dengan menghitung indeks Davies-Bouldin menggunakan `davies_bouldin_score()`. Nilai indeks Davies-Bouldin yang lebih rendah menunjukkan bahwa clustering yang dihasilkan lebih baik, karena mengindikasikan bahwa cluster lebih terpisah dan lebih padat.

#### E. Visualisasi Hasil

```
# --- 4. Visualizing the Results ---
# Visualisasi hasil clustering
```

```

plt.figure(figsize=(8, 6))
plt.scatter(reduced_data[:, 0], reduced_data[:, 1], c=kmeans.labels_,
            cmap='viridis')
plt.title('Hasil Clustering Kampanye Berdasarkan Karakteristik Audiens')
plt.xlabel('PCA 1')
plt.ylabel('PCA 2')
plt.colorbar(label='Cluster')
plt.show()

# Rangkuman jumlah data per cluster
print("Jumlah data per cluster:")
cluster_counts = df_selected['Cluster'].value_counts().sort_index()
print(cluster_counts)

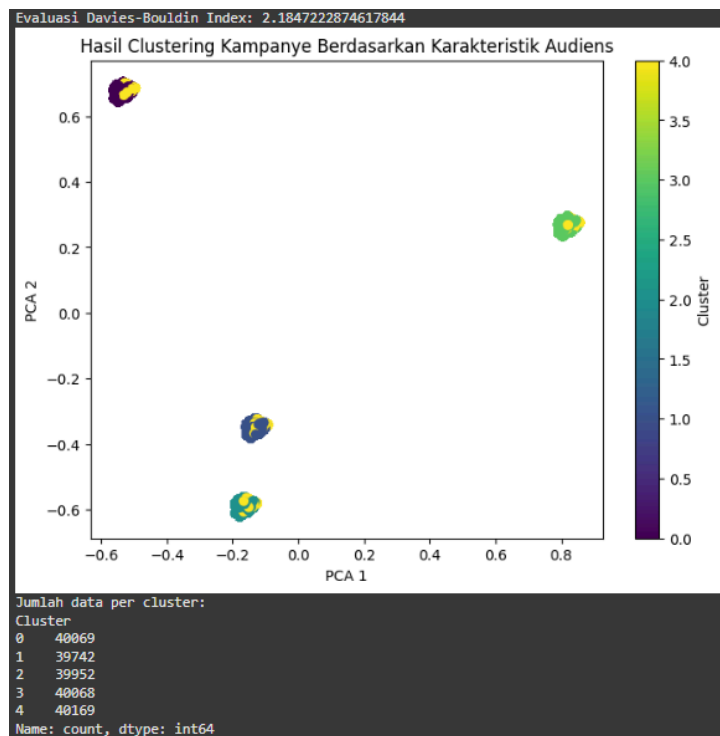
# Menampilkan ringkasan nilai terbanyak (modus) dari setiap cluster
print("\nRangkuman nilai terbanyak (modus) dari setiap cluster:")
for cluster_id in sorted(df_selected['Cluster'].unique()):
    cluster_data = df_selected[df_selected['Cluster'] == cluster_id]
    print(f"\nCluster {cluster_id}:")
    print("Jumlah Data:", len(cluster_data))
    print("Nilai Terbanyak (Modus) untuk Setiap Kolom:")
    print(cluster_data.mode().iloc[0]) # Menampilkan modus dari setiap kolom

```

Gambar 4.13 Code Visualisasi Hasil

Pada Gambar 4.13, kode ini memvisualisasi hasil clustering menggunakan PCA dilakukan untuk mereduksi dimensi data sehingga dapat divisualisasikan dalam dua dimensi. `reduced_data[:, 0]` dan `reduced_data[:, 1]` adalah dua komponen utama yang dihasilkan dari PCA dan digunakan sebagai sumbu X dan Y dalam plot scatter. Setiap titik pada scatter plot mewakili sebuah data yang dikelompokkan ke dalam salah satu cluster yang dihasilkan oleh algoritma K-means, dengan warna berbeda untuk masing-masing cluster, yang diindikasikan oleh `kmeans.labels_`. Judul plot memberikan konteks bahwa ini adalah hasil clustering kampanye berdasarkan karakteristik audiens. Sumbu X dan Y diberi label "PCA 1" dan "PCA 2" untuk menggambarkan komponen utama pertama dan kedua dari hasil reduksi dimensi. Kemudian, setelah visualisasi, kode ini menampilkan jumlah data yang ada di masing-masing cluster dengan `value_counts()` dan memberikan informasi tentang nilai terbanyak (modus) untuk setiap kolom dalam setiap cluster, yang memberikan wawasan tentang karakteristik dominan dari tiap cluster.

## F. Output Code Bagian 2f



Gambar 4.14 Output Bag.1 Code K-Means Kampanye Marketing

```
Cluster 0:
Jumlah Data: 40069
Nilai Terbanyak (Modus) untuk Setiap Kolom:
Campaign_Type      Search
Target_Audience   Men 18-24
Duration            45 days
Channel_Used       YouTube
Cluster            0
Name: 0, dtype: object

Cluster 1:
Jumlah Data: 39742
Nilai Terbanyak (Modus) untuk Setiap Kolom:
Campaign_Type      Display
Target_Audience   All Ages
Duration            15 days
Channel_Used       Email
Cluster            1
Name: 0, dtype: object

Cluster 2:
Jumlah Data: 39952
Nilai Terbanyak (Modus) untuk Setiap Kolom:
Campaign_Type      Search
Target_Audience   Women 35-44
Duration            60 days
Channel_Used       YouTube
Cluster            2
Name: 0, dtype: object

Cluster 3:
Jumlah Data: 40068
Nilai Terbanyak (Modus) untuk Setiap Kolom:
Campaign_Type      Search
Target_Audience   Women 25-34
Duration            30 days
Channel_Used       Website
Cluster            3
Name: 0, dtype: object

Cluster 4:
Jumlah Data: 40169
Nilai Terbanyak (Modus) untuk Setiap Kolom:
Campaign_Type      Influencer
Target_Audience   All Ages
Duration            30 days
Channel_Used       YouTube
Cluster            4
Name: 0, dtype: object
```

Gambar 4.15 Output Bag.2 Code K-Means Kampanye Marketing

Dari Gambar 4.14 dan Gambar 4.15 Hasil clustering membagi data menjadi lima kelompok dengan distribusi yang relatif seimbang, di mana Cluster 0 berisi 40.069 data dengan karakteristik dominan kampanye Search, target audiens Men 18-24, durasi 45 days, dan saluran YouTube. Cluster 1, dengan 39.742 data, didominasi oleh kampanye Display, target audiens All Ages, durasi 15 days, dan saluran Email. Cluster 2 memiliki 39.952 data dengan kampanye Search, target audiens Women 35-44, durasi 60 days, dan saluran YouTube. Cluster 3 mencakup 40.068 data dengan kampanye Search, target audiens Women 25-34, durasi 30 days, dan saluran Website. Terakhir, Cluster 4 dengan 40.169 data didominasi oleh kampanye Influencer, target audiens All Ages, durasi 30 days, dan saluran YouTube. Rangkuman ini menunjukkan pola yang berbeda pada setiap cluster, dengan perbedaan mencolok pada jenis kampanye, target audiens, durasi, dan saluran yang digunakan.

Selain itu, dilakukan evaluasi model clustering menggunakan Davies-Bouldin Index (DB Index). Semakin kecil nilai DB Index, maka semakin baik kualitas clustering yang dihasilkan. Pada output, nilai DB Index yang diperoleh adalah 2.1847222874617844, menunjukkan bahwa clustering yang dihasilkan cukup baik.

## **BAB V**

### **KESIMPULAN DAN SARAN**

Pendekatan menggunakan algoritma K-Means Clustering untuk segmentasi kampanye marketing berdasarkan karakteristik audiens mampu memberikan hasil segmentasi yang jelas. Algoritma K-Means dapat mengidentifikasi dan memisahkan kelompok data dengan karakteristik yang berbeda. Setiap cluster yang dihasilkan memiliki karakteristik yang jelas meliputi jenis kampanye, target audiens, durasi, dan saluran yang digunakan. Penggunaan Davies Bouldin Index sebagai evaluasi menunjukkan nilai yang relatif kecil, yaitu 2.1847222874617844, yang menandakan bahwa clustering yang dihasilkan cukup baik. Semakin kecil nilai DBI, semakin baik pemisahan antar cluster, artinya algoritma K-Means clustering berhasil membedakan kelompok data dengan baik.

Untuk hasil segmentasi yang lebih efektif, disarankan untuk melakukan analisis lebih mendalam terhadap variabel-variabel yang mempengaruhi efektivitas kampanye pemasaran, seperti perilaku konsumen dan tren pasar terkini. Selain itu, menerapkan algoritma pembelajaran mesin yang lebih canggih, seperti algoritma Density-Based Spatial Clustering of Application with Noise (DBSCAN), Deep Learning, Latent Dirichlet Allocation (LDA), atau Gaussian Mixture Model (GMM) dapat memberikan hasil yang lebih akurat dan relevan. Pilihan algoritma dapat disesuaikan tergantung karakteristik dataset. Evaluasi hasil clustering juga dapat diperluas dengan menggunakan metrik tambahan, seperti Silhouette Score, untuk memastikan kualitas segmentasi yang lebih baik.

## DAFTAR PUSTAKA

- [1] Erhard Rahm, H. Do, Data Cleaning: Problems and Current Approaches, IEEE Data Engineering Bulletin, 2000, Vol.4 pp.3–13.
- [2] E. A. Wibowo and S. L. Ratnasari, “OPTIMIZING THE EFFECTIVENESS OF DIGITAL MARKETING CAMPAIGNS WITH THE USE OF E-COMMERCE COMPANY DECISION SUPPORT SYSTEMS”, *ICOBUSS*, vol. 4, no. 1, pp. 284–292, Nov. 2024, doi: <https://doi.org/10.24034/icobuss.v4i1.505>
- [3] I. Bagus, Retantyo Wardoyo, and A. K. Sari, “The K-Means Clustering Algorithm With Semantic Similarity To Estimate The Cost of Hospitalization,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 13, no. 4, pp. 313–322, 2019, Accessed: Dec. 08, 2024. [Online]. Available: <https://jurnal.ugm.ac.id/ijccs/article/view/45093/26039>
- [4] F. Ros, R. Riad, and S. Guillaume, “PDBI: A partitioning Davies-Bouldin index for clustering evaluation,” *Neurocomputing*, vol. 528, pp. 178–199, Apr. 2023, doi: 10.1016/j.neucom.2023.01.043.
- [5] A. R. Afandi and H. Kurnia, “Academy of Social Science and Global Citizenship Journal Revolusi Teknologi: Masa Depan Kecerdasan Buatan (AI) dan Dampaknya Terhadap Masyarakat,” *AoSSaGCJ*, vol. 3, no. 1, 2023, doi: 10.47200/AoSSaGCJ.v3i1.1837.
- [6] W. El-Shafai *et al.*, “Hybrid Segmentation Approach for Different Medical Image Modalities,” *Computers, Materials & Continua*, vol. 73, no. 2, pp. 3455–3472, 2022, doi: 10.32604/cmc.2022.028722.
- [7] L. Rahmawati, J. Informatika, S. W. Sihwi, and E. Suryani, “ANALISA CLUSTERING MENGGUNAKAN METODE K-MEANS DAN HIERARCHICAL CLUSTERING (STUDI KASUS: DOKUMEN SKRIPSI JURUSAN KIMIA, FMIPA, UNIVERSITAS SEBELAS MARET).”
- [8] Y. Hasan, “PENGUKURAN SILHOUETTE SCORE DAN DAVIES-BOULDIN INDEX PADA HASIL CLUSTER K-MEANS DAN DBSCAN,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Oct. 2024, doi: 10.23960/jitet.v12i3S1.5001.
- [9] D. N. P. Sari, “ANALISIS CLUSTER DENGAN METODE K-MEANS PADA PERSEBARAN KASUS COVID-19 BERDASARKAN PROVINSI DI INDONESIA,” Dokumen Akademik Online, UNIVERSITAS NEGERI SEMARANG, 2020. Accessed: Dec. 04, 2024. [Online]. Available: <https://lib.unnes.ac.id/41838/1/4112317029.pdf>



## LAMPIRAN

- Laporan kelompok 9 K-Means Clustering  
<https://docs.google.com/document/d/1zO4V7d-euCEuWXHuNhkhHo9c3bWPegXRjZH-IIeU9wc/edit?usp=sharing>
- Dataset Marketing Campaign  
<https://www.kaggle.com/datasets/manishabhattach22/marketing-campaign-performance-data-set>
- Google Collab  
[https://colab.research.google.com/drive/1\\_6MVbR93dqvYRn3XUJU3AtdFE38f7Ycs?usp=sharing](https://colab.research.google.com/drive/1_6MVbR93dqvYRn3XUJU3AtdFE38f7Ycs?usp=sharing)
- Form Asistensi  
[https://drive.google.com/drive/folders/1B\\_IzMz833fVgsIzq2uIUzy0Pn4CT2NIH?usp=sharing](https://drive.google.com/drive/folders/1B_IzMz833fVgsIzq2uIUzy0Pn4CT2NIH?usp=sharing)