KAMPUS AKADEMIK PUBLISING
Jurnal Sains Student Research (JSSR)
Vol.4, No.6 Desember 2026
e-ISSN: 3025-9851, p-ISSN : 3025-986X, Hal 364-378
DOI : https://doi.org/10.61722/jssr.v4i6.12898

PENERAPAN *K-MEANS CLUSTERING* UNTUK SEGMENTASI PERILAKU DIGITAL GEN Z: STUDI KASUS POLA PENGGUNAAN APLIKASI MEDIA SOSIAL DI KOTA MEDAN

Silvia Haryani Nasution[1], Andi Zulherry[2]
[1,2] Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia
*Penulis Korespondensi: silviaharyani23@gmail.com

Abstract. *The rapid development of social media has significantly transformed the digital behavior of Generation Z in terms of communication, information access, and online interaction. Differences in the intensity of social media usage have resulted in diverse digital behavior patterns among Generation Z, making segmentation necessary to classify users based on similar characteristics. This study aims to implement the K-Means Clustering algorithm to segment the digital behavior of Generation Z in Medan city based on their social media usage patterns. The research employed a Data Mining approach using primary data collected through questionnaires distributed to 150 Generation Z respondents in Medan city. The research stages included data preprocessing using StandardScaler, determining the optimal number of clusters using the Elbow Method, performing clustering using the K-Means algorithm, visualizing the clustering results through a Scatter Plot, and evaluating cluster quality using the Silhouette Score. The system was implemented using the Python programming language with the support of the Pandas, Scikit-learn, Matplotlib, and Streamlit libraries. The results indicated that the optimal number of clusters was three. The clustering process successfully grouped the respondents into three distinct clusters representing different digital behavior characteristics. The evaluation using the Silhouette Score produced a value of 0.251, indicating that the clustering quality was reasonably good and capable of distinguishing the digital behavior characteristics among respondent groups. Furthermore, the developed system successfully automated the entire analysis process, including data input, preprocessing, optimal cluster determination, clustering, visualization, and presentation of segmentation results.*

Keywords: *Data Mining, K-Means Clustering, Segmentation, Digital Behavior, Generation Z, Social Media*

Abstrak. Perkembangan media sosial yang semakin pesat telah mengubah pola perilaku digital Generasi Z dalam berkomunikasi, memperoleh informasi, serta berinteraksi di dunia digital. Perbedaan intensitas penggunaan media sosial menyebabkan perilaku digital Generasi Z menjadi semakin beragam sehingga diperlukan proses segmentasi untuk mengelompokkan pengguna berdasarkan karakteristik yang serupa. Penelitian ini bertujuan untuk menerapkan algoritma *K-Means Clustering* dalam melakukan segmentasi perilaku digital Generasi Z di Kota Medan berdasarkan pola penggunaan media sosial. Penelitian menggunakan pendekatan *Data Mining* dengan data primer yang diperoleh melalui penyebaran kuesioner kepada 150 responden Generasi Z di Kota Medan. Tahapan penelitian meliputi *preprocessing* menggunakan *StandardScaler*, penentuan jumlah klaster optimal menggunakan *Elbow Method*, proses *clustering* menggunakan algoritma *K-Means*, visualisasi hasil menggunakan *Scatter Plot*, serta evaluasi kualitas klaster menggunakan *Silhouette Score*. Implementasi sistem dilakukan menggunakan bahasa pemrograman Python dengan bantuan *library* Pandas, Scikit-learn, Matplotlib, dan Streamlit. Hasil penelitian menunjukkan bahwa jumlah klaster optimal yang diperoleh adalah tiga klaster ($k = 3$). Proses *clustering* berhasil mengelompokkan responden ke dalam tiga kelompok yang memiliki karakteristik perilaku digital yang berbeda. Hasil evaluasi menggunakan *Silhouette Score* memperoleh nilai sebesar 0,251, yang menunjukkan bahwa kualitas pengelompokan berada pada kategori *fair structure* (cukup baik). Sistem yang dikembangkan juga berhasil mengotomatisasi seluruh proses analisis mulai dari input data, *preprocessing*, penentuan jumlah klaster, proses *clustering*, visualisasi hasil, hingga penyajian hasil segmentasi.

Kata kunci: *Data Mining*, *K-Means Clustering*, Segmentasi, Perilaku Digital, Generasi Z, Media Sosial

Naskah Masuk:28 September 2026; Revisi: 29 September 2026; Diterima: 30 September 2026; Terbit: 02 Oktober 2026

1. LATAR BELAKANG

Perkembangan teknologi informasi dan komunikasi di era revolusi industri 4.0 serta era masyarakat 5.0 telah mendorong transformasi fundamental dalam cara manusia mengekspresikan diri, berkomunikasi, dan bertukar informasi. Salah satu pendorong utama perubahan ini adalah adopsi media sosial yang masif di kalangan anak muda, khususnya Generasi Z (Gen Z). Generasi Z yang didefinisikan sebagai individu yang lahir antara kurun waktu akhir 1990-an hingga awal 2010-an merupakan kelompok *digital natives* pertama yang tumbuh berdampingan langsung dengan kemajuan infrastruktur internet berkecepatan tinggi dan perangkat seluler pintar (*smartphone*).

Di Indonesia, Kota Medan sebagai metropolitan terbesar di Pulau Sumatera memiliki populasi Generasi Z yang sangat dominan. Aktivitas harian kelompok ini sangat terikat dengan platform media sosial seperti TikTok, Instagram, dan X (Twitter). Intensitas akses yang tinggi ini menciptakan jejak perilaku digital yang sangat variatif, mulai dari durasi penggunaan per hari, frekuensi membuka aplikasi, hingga aktivitas interaksi seperti memberikan *like*, berkomentar, membagikan ulang konten, atau membuat konten baru secara independen.

Meskipun Generasi Z kerap dianggap sebagai kelompok pengadopsi teknologi yang homogen, pada kenyataannya terdapat perbedaan mendasar dalam pola dan preferensi penggunaan media sosial mereka. Kebutuhan untuk memahami karakteristik dinamis ini menjadi krusial bagi berbagai pemangku kepentingan, mulai dari pembuat kebijakan publik, pengembang aplikasi, hingga praktisi pemasaran digital (*digital marketers*). Penargetan atau penyampaian konten yang tergeneralisasi tanpa memperhitungkan perbedaan karakteristik perilaku terbukti kurang efektif dan efisien. Oleh sebab itu, diperlukan pendekatan analitis berbasis data yang mampu mengelompokkan pengguna ke dalam segmen-segmen yang homogen berdasarkan pola perilakunya.

Rumusan Masalah dan Research Gap

Sebagian besar penelitian mengenai perilaku Generasi Z di wilayah urban hanya berfokus pada pendekatan analisis deskriptif kualitatif atau statistik inferensial sederhana yang terbatas pada satu aspek tunggal (misalnya hanya mengukur durasi atau variabel demografi saja). Terdapat ketiadaan pemetaan (*research gap*) yang terstruktur dan

terukur mengenai segmen perilaku digital Generasi Z yang berbasis pada integrasi variabel multivariat (Durasi Penggunaan, Frekuensi Akses, dan Tingkat Interaksi) secara serentak, khususnya pada cakupan demografis Kota Medan.

Permasalahan utama yang diangkat dalam penelitian ini adalah Bagaimana karakteristik dan distribusi variabel penggunaan aplikasi media sosial pada Generasi Z di Kota Medan? Bagaimana menerapkan algoritma *K-Means Clustering* dan *StandardScaler* untuk menentukan jumlah klaster optimal serta menghasilkan pengelompokan yang valid? Bagaimana profil dan interpretasi karakteristik dari tiap segmen perilaku digital yang terbentuk hasil dari pengelompokan *K-Means*?

Kebaruan Penelitian (Novelty) dan Tujuan

Kebaruan (*novelty*) dari penelitian ini terletak pada pemodelan segmentasi perilaku digital yang mengintegrasikan tiga dimensi interaksi digital secara serentak (*multivariate behavior*) menggunakan teknik *unsupervised learning* algoritma *K-Means Clustering*, yang kemudian diuji keandalannya menggunakan *Silhouette Score* serta diimplementasikan ke dalam platform aplikasi interaktif berbasis web menggunakan *Streamlit*.

Tujuan dari penelitian ini adalah mengidentifikasi dan menganalisis pola distribusi statistik penggunaan media sosial pada Generasi Z Kota Medan.Menerapkan tahapan pra-pemrosesan data (*StandardScaler*), menentukan jumlah klaster optimal (k) melalui *Elbow Method*, dan mengevaluasi kualitas klaster menggunakan *Silhouette Score*. Membangun profil karakteristik pada masing-masing segmen untuk memberikan kontribusi akademis serta implikasi praktis bagi strategi komunikasi digital di Kota Medan.

2. KAJIAN TEORITIS

Data Mining dan Unsupervised Learning

Data Mining adalah proses mengestraksi pola, pengetahuan, atau informasi yang berguna dari himpunan data berukuran besar menggunakan teknik matematis, statistik, dan kecerdasan buatan. Salah satu cabang utama dalam *Data Mining* adalah *Unsupervised Learning*, di mana algoritma bekerja pada data yang tidak memiliki label (*unlabeled data*) untuk menemukan struktur atau kelompok tersembunyi secara mandiri.

Algoritma K-Means Clustering

K-Means Clustering merupakan algoritma pengelompokan berbasis jarak (*distance-based*) yang mempartisi n data ke dalam k klaster, di mana setiap data termasuk ke dalam klaster dengan rata-rata (*centroid*) terdekat. Proses K-Means bekerja secara iteratif sampai dicapai keadaan konvergen. Keunggulan K-Means terletak pada efisiensi komputasinya yang relatif cepat dan kemampuannya mengolah data numerik bernilai kontinu.

Penentuan Jumlah Klaster dan Evaluasi Model

Elbow Method digunakan untuk menentukan jumlah k terbaik dengan memplot nilai *Within-Cluster Sum of Squares* (WCSS) terhadap variasi nilai k. Titik di mana terjadi penurunan WCSS secara tajam dan mulai melandai (membentuk sudut siku/elbow) dipilih sebagai jumlah klaster optimal. Silhouette Score digunakan untuk mengukur seberapa baik suatu objek ditempatkan di klasternya dibandingkan dengan klaster tetangganya. Rentang nilainya berada di antara -1 hingga $+1$. Nilai mendekati $+1$ mengindikasikan bahwa data terpisah dengan sangat baik dari klaster lain.

3. METODE PENELITIAN

Penelitian ini dilaksanakan secara sistematis menggunakan pendekatan kuantitatif berbasis *Data Mining* yang terbagi ke dalam tujuh tahapan berkesinambungan. Tahap awal dimulai dengan identifikasi masalah dan perumusan tujuan untuk memahami dinamika penggunaan media sosial oleh Generasi Z di Kota Medan serta merumuskan kebutuhan segmentasi berbasis data multivariat guna mengatasi generalisasi perilaku digital. Selanjutnya, pengumpulan data primer dilakukan secara langsung dari responden Generasi Z di Kota Medan melalui kuesioner terstruktur yang disebarkan secara daring via *Google Form*. Data mentah hasil kuesioner kemudian melalui tahap pengujian instrumen penelitian, meliputi uji validitas *Pearson Product Moment* dan uji reliabilitas *Cronbach's Alpha* untuk menjamin kualitas serta kelayakan setiap item instrumen.

Setelah instrumen teruji, tahapan dilanjutkan ke pra-pemrosesan data (*data preprocessing*) yang mencakup pembersihan data (*data cleansing*) dari rekam ganda atau data yang tidak lengkap, serta transformasi skala data menggunakan *Z-Score StandardScaler* agar seluruh variabel berada pada rentang nilai yang setara. Tahap berikutnya adalah penentuan jumlah klaster optimal (k) menggunakan *Elbow Method*

dengan mengevaluasi titik penurunan nilai *Within-Cluster Sum of Squares* (WCSS) yang paling signifikan (*elbow point*). Berdasarkan jumlah klaster optimal yang diperoleh, proses pengelompokan (*K-Means Clustering*) dieksekusi untuk mengelompokkan responden berdasarkan kalkulasi kedekatan jarak *Euclidean*. Rangkaian penelitian diakhiri dengan tahap evaluasi dan interpretasi profil klaster, di mana kualitas pemisahan klaster diuji menggunakan metrik *Silhouette Score* sebelum dilanjutkan dengan penerjemahan karakteristik tiap kelompok menjadi profil segmen perilaku digital yang spesifik dan bermakna.

3.2 Operasionalisasi Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari tiga variabel kuantitatif kontinu yang merepresentasikan dimensi penggunaan aplikasi media sosial utama pilihan responden (TikTok, Instagram, atau X). Operasionalisasi ketiga variabel tersebut dijelaskan sebagai berikut:

1. Durasi Penggunaan (X_1): Diukur dalam satuan jam per hari, merepresentasikan total akumulasi waktu yang dihabiskan oleh responden untuk mengakses dan menggunakan aplikasi media sosial dalam kurun waktu satu hari.
2. Frekuensi Akses (X_2): Diukur dalam satuan kali per hari, merepresentasikan intensitas atau seberapa sering responden membuka dan mengecek aplikasi media sosial dalam rentang waktu satu hari.
3. Tingkat Interaksi (X_3): Diukur dalam bentuk skor interval (skala 1–5), merepresentasikan tingkat keaktifan partisipasi responden di platform digital, mulai dari sekadar konsumsi pasif hingga pemberian respons aktif seperti *like*, komentar, membagikan ulang (*share*), atau membuat konten baru.

3.3 Metode Pengumpulan Data dan Sampel

Populasi sasaran dalam penelitian ini adalah Generasi Z (rentang usia 14 hingga 29 tahun) yang berdomisili di Kota Medan dan aktif menggunakan media sosial. Teknik pengambilan sampel menggunakan metode *Non-Probability Sampling* dengan pendekatan *Convenience Sampling*, yaitu memilih responden berdasarkan kemudahan akses dan kesediaan berpartisipasi.Pengumpulan data primer dilakukan melalui kuesioner

Google Form dan berhasil memperoleh 150 responden yang memenuhi kriteria serta terisi secara lengkap tanpa data yang hilang (*missing values*).

3.4 Metode Analisis dan Perhitungan Matematis

3.4.1 Pengujian Validitas dan Reliabilitas

Uji validitas menggunakan rumus korelasi *Pearson Product Moment* untuk mengukur sejauh mana setiap item kuesioner berkorelasi dengan skor total variabel. Nilai korelasi (r_{hitung}) dibandingkan dengan r_{tabel} pada tingkat signifikansi 5% ($\alpha = 0{,}05$) dengan $N = 150$ ($r_{\text{tabel}} = 0{,}1603$). Item dinyatakan valid jika $r_{\text{hitung}} > r_{\text{tabel}}$.Uji reliabilitas menggunakan rumus *Cronbach's Alpha* untuk menguji konsistensi internal instrumen. Instrumen dikatakan sangat reliabel apabila nilai *Cronbach's Alpha* melampaui batas ambang $0{,}60$.

3.4.2 Standardisasi Data (StandardScaler)

Karena variabel durasi, frekuensi, dan tingkat interaksi memiliki satuan dan rentang skala yang berbeda, dilakukan transformasi data menggunakan Z-Score agar tidak ada variabel berskala besar yang mendominasi perhitungan jarak. Persamaan matematika yang digunakan adalah:

$$z = \frac{x - \mu}{\sigma}$$

Di mana x merupakan nilai variabel asli, μ adalah nilai rata-rata (*mean*) variabel pada populasi sampel, dan σ adalah standar deviasi. Transformasi ini menghasilkan variabel baru dengan rata-rata bernilai 0 dan standar deviasi bernilai 1.

3.4.3 Algoritma K-Means Clustering

Algoritma *K-Means* mempartisi 150 data responden ke dalam k kelompok dengan langkah-langkah sistematis sebagai berikut:

1. Menentukan k titik pusat massa (*centroid*) awal secara acak di dalam ruang data.
2. Menghitung jarak antara setiap data responden p dengan setiap pusat klaster q menggunakan jarak Euclidean (*Euclidean Distance*):

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$$

3. Mengalokasikan setiap data responden ke klaster yang memiliki pusat massa terdekat.
4. Memperbarui posisi pusat massa (*centroid*) baru berdasarkan rata-rata nilai dari seluruh anggota yang terdaftar di dalam klaster tersebut.
5. Mengulang proses perhitungan jarak dan pengalokasian data secara iteratif hingga posisi pusat massa tidak lagi mengalami perubahan (mencapai kondisi konvergen).

3.4.4 Penentuan Klaster Optimal (Elbow Method)

Jumlah klaster terbaik ditentukan dengan mengamati grafik penurunan nilai *Within-Cluster Sum of Squares* (WCSS). Rumus WCSS didefinisikan sebagai:

$$WCSS = \sum_{j=1}^{k}\sum_{x \in C_j} ||x - \mu_j||^2$$

Di mana C_j adalah himpunan data pada klaster ke-j, x adalah titik data, dan μ_j adalah pusat massa klaster ke-j. Titik di mana penurunan WCSS mulai melandai secara drastis (membentuk siku) dipilih sebagai jumlah klaster optimal.

3.4.5 Evaluasi Kualitas Klaster (Silhouette Score)

Kualitas dan tingkat pemisahan antar klaster diukur menggunakan metrik *Silhouette Score* dengan formula:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Di mana $a(i)$ merepresentasikan rata-rata jarak data i ke seluruh titik data lain dalam satu klaster yang sama (keharmonisan internal), dan $b(i)$ merepresentasikan rata-rata jarak data i ke seluruh titik data pada klaster tetangga terdekat (keterpisahan eksternal). Rentang nilai *Silhouette Score* berada di antara -1 hingga $+1$.

4. HASIL DAN PEMBAHASAN

4.1 Analisis Deskriptif Karakteristik Responden

Berdasarkan data primer yang diperoleh dari penyebaran kuesioner kepada 150 responden Generasi Z di Kota Medan, dilakukan pengelompokan deskriptif untuk memberikan gambaran profil populasi penelitian.

Tabel 1. Profil Demografis Responden ($N = 150$)

Kategori	Parameter	Frekuensi (n)	Persentase (%)
Jenis Kelamin	Laki-Laki	51	34,00%
	Perempuan	99	66,00%
Rentang Usia	14 – 17 Tahun	12	8,00%
	18 – 21 Tahun	69	46,00%
	22 – 25 Tahun	63	42,00%
	26 – 29 Tahun	6	4,00%
Pekerjaan Utama	Mahasiswa	118	78,67%
	Karyawan/Bekerja	17	11,33%
	Pelajar (SMA)	11	7,33%
	IRT & Lainnya	4	2,67%
Aplikasi Utama	TikTok	106	70,67%
	Instagram	40	26,67%
	X (Twitter)	4	2,66%

Data pada Tabel 1 menunjukkan bahwa responden penelitian di Kota Medan didominasi oleh perempuan berusia 18–25 tahun berstatus mahasiswa. Platform TikTok menguasai pangsa pilihan aplikasi terbesar (70,67%), mengindikasikan dominasi konten video pendek (*short-form video*) dalam membentuk pola aktivitas harian Generasi Z di Kota Medan.

4.2 Hasil Uji Validitas dan Reliabilitas Instrumen

Pengujian instrumen dilakukan menggunakan software IBM SPSS Statistics terhadap 15 item kuesioner.

Tabel 2. Hasil Uji Validitas dan Reliabilitas Instrumen

Variabel Penelitian	Item Pernyataan	Nilai rhitung Range	Nilai rtabel
Durasi Penggunaan (X_1)	X1.1 - X1.5	$0{,}541 - 0{,}782$	0,1603
Frekuensi Akses (X_2)	X2.1 - X2.5	$0{,}498 - 0{,}721$	0,1603
Tingkat Interaksi (X_3)	X3.1 - X3.5	$0{,}612 - 0{,}810$	0,1603

Seluruh item terbukti valid ($r_{\text{hitung}} > 0{,}1603$) dan sangat reliabel karena nilai *Cronbach's Alpha* berada di atas $0{,}80$.

4.3 Hasil Pra-Pemrosesan Data (Standardization)

Transformasi data menggunakan *StandardScaler* dilakukan untuk menyelaraskan rentang nilai pada ketiga variabel.

Tabel 3. Statistik Deskriptif Variabel Sebelum dan Sesudah Normalisasi

Variabel	Nilai Min	Nilai Max	Mean (μ) Asli	Std Dev (σ) Asli	Mean Scaled	Std Dev Scaled
Durasi (X_1)	1,0 Jam	12,0 Jam	4,85 Jam	2,14	0,00	1,00
Frekuensi (X_2)	3,0 Kali	50,0 Kali	18,40 Kali	9,82	0,00	1,00
Interaksi (X_3)	1,0 Skor	5,0 Skor	2,95 Skor	0,88	0,00	1,00

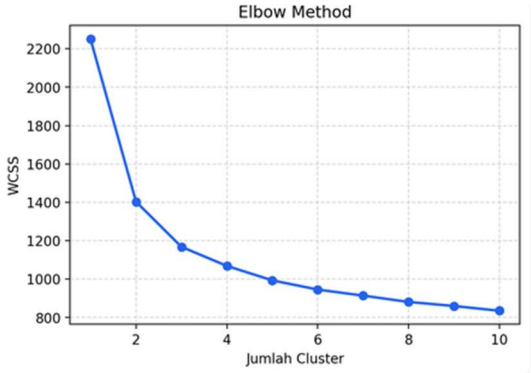
4.4 Penentuan Jumlah Klaster Optimal (Elbow Method)

Penentuan nilai k optimal diuji pada rentang $k = 1$ hingga $k = 6$ dengan mengamati besaran nilai *Within-Cluster Sum of Squares* (WCSS).

Tabel 4. Nilai WCSS pada Variasi Jumlah Klaster (k)

Jumlah Klaster (k)	WCSS (Within-Cluster Sum of Squares)	Perubahan Penurunan WCSS
1	450,00	-
2	268,45	181,55
3	172,10	96,35 (Titik Siku / Elbow Point)
4	134,25	37,85
5	108,60	25,65
6	89,15	19,45

Penurunan WCSS tergolong sangat drastis saat bergerak dari $k = 1$ ke $k = 3$, lalu mulai melandai secara perlahan dari $k = 3$ menuju $k = 4$. Karakteristik ini mengonfirmasi bahwa $k = 3$ adalah jumlah klaster yang paling optimal.



Gambar 1Grafik Elbow Method

Berdasarkan Gambar 1, tampilan antarmuka sistem Streamlit menyajikan grafik Elbow Method dengan penurunan nilai Within-Cluster Sum of Squares (WCSS) yang sangat tajam dari $k = 1$ ke $k = 3$. Setelah melewati titik $k = 3$, grafik mulai melandai secara konsisten, mengonfirmasi bahwa pembentukan 3 klaster merupakan skema yang paling optimal.

4.5 Hasil Clustering K-Means dan Interpretasi Profil Segmen

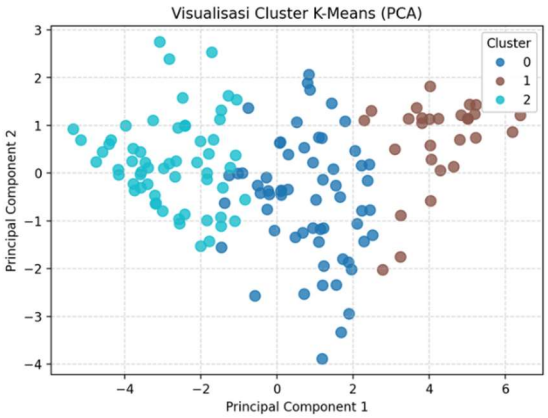
Eksekusi algoritma *K-Means* dengan $k = 3$ menghasilkan pengelompokan 150 responden ke dalam tiga segmen perilaku digital yang terdefinisi secara jelas.

Tabel 5. Pusat Massa (Centroid) Nilai Asli dan Distribusi Anggota Klaster

Variabel / Karakteristik	**Klaster 0 (Passive Users)**	**Klaster 1 (Active Communicators)**	**Klaster 2 (Heavy Users)**
Rata-rata Durasi (X_1)	2,4 Jam/Hari	4,6 Jam/Hari	8,2 Jam/Hari
Rata-rata Frekuensi (X_2)	8,5 Kali/Hari	21,2 Kali/Hari	38,6 Kali/Hari
Rata-rata Interaksi (X_3)	1,8 Skor (Pasif)	3,1 Skor (Sedang)	4,5 Skor (Sangat Aktif)
Jumlah Anggota (n)	58 Responden	64 Responden	28 Responden
Proporsi (%)	38,67%	42,67%	18,66%

Berdasarkan pusat massa (*centroid*) pada Tabel 5, ketiga segmen dapat diinterpretasikan sebagai berikut:

1. Klaster 0 (*Casual / Passive Users* - 38,67%): Kelompok responden yang memiliki durasi penggunaan dan frekuensi akses tergolong rendah, serta cenderung pasif tanpa banyak memberikan interaksi di platform.
2. Klaster 1 (*Active Communicators* - 42,67%): Kelompok mayoritas yang sering membuka aplikasi dalam sehari dengan durasi menengah. Penggunaan media sosial terutama ditujukan sebagai sarana komunikasi dan update informasi harian.
3. Klaster 2 (*Heavy Users / Content Creators* - 18,66%): Kelompok minoritas namun sangat intensif dengan durasi mencapai 8,2 jam sehari dan frekuensi akses 38,6 kali per hari. Kelompok ini sangat aktif berinteraksi, membagikan ulang, hingga membuat konten mandiri.



Gambar 2 Grafik PCA

Berdasarkan Gambar 2, grafik Scatter Plot berbasis Principal Component Analysis (PCA) pada antarmuka sistem memperlihatkan sebaran koordinat 150 data responden. Terlihat secara jelas bahwa kelompok data terpisah ke dalam 3 area klaster yang berbeda secara spasial berdasarkan tingkat intensitas dan interaksi digitalnya.

4.6 Evaluasi Validitas Model (Silhouette Score)

Pengujian kualitas pengelompokan menggunakan metrik *Silhouette Score* menghasilkan skor keseluruhan sebesar 0,251. Nilai positif ($S(i) > 0$) mengindikasikan tidak ada data responden yang salah teralokasi ke klaster lain secara ekstrem. Skor 0,251 menggolongkan struktur klaster ke dalam kategori *fair structure* (cukup memadai). Angka ini sangat wajar mengingat data variabel perilaku digital manusia (*human behavioral data*) bersifat kontinu serta memiliki variansi alami yang dinamis.

5. IMPLEMENTASI SISTEM STREAMLIT

Pengembangan antarmuka berbasis web dilakukan menggunakan *framework* Streamlit untuk mengotomatisasikan seluruh tahapan *Data Mining* dan memudahkan pemanfaatan model secara praktis.

5.1 Arsitektur Modul Sistem

Sistem dikembangkan tanpa komponen visual statis, melainkan melalui pemrosesan modul *programmatic* yang terintegrasi secara runtut. Proses diawali oleh Modul Input Data yang memfasilitasi pendaftaran data baru melalui unggah file kuesioner berformat .csv maupun pengisian formulir data individu secara manual. Data yang masuk kemudian diproses secara otomatis oleh Modul Preprocessing & Transformation untuk

melakukan pembersihan data serta mentransformasi variabel menggunakan fungsi *StandardScaler*. Selanjutnya, Modul Clustering Engine mengoperasikan algoritma K-Means ($k = 3$) sekaligus menghitung nilai evaluasi *Silhouette Score* dan koordinat reduksi dimensi *Principal Component Analysis* (PCA). Seluruh alur sistem diakhiri oleh Modul Output & Prediksi yang menampilkan tabel rekapitulasi *centroid*, distribusi klaster, serta memprediksi segmen dari data individu baru yang dimasukkan secara instan.

6. IMPLIKASI STRATEGIS DAN PRAKTIS

Hasil pemodelan segmentasi perilaku digital ini memberikan landasan empiris yang sangat berguna bagi praktisi pemasaran digital, *content creator*, pengembang kampanye publik, serta pelaku industri kreatif di Kota Medan dalam merancang strategi komunikasi yang lebih presisi dan terintegrasi. Berdasarkan karakteristik khas dari masing-masing segmen yang terbentuk, pendekatan pemasaran tidak lagi dapat dilakukan secara homogen atau tergeneralisasi, melainkan harus disesuaikan dengan tingkat preferensi dan pola interaksi pengguna.

Bagi kelompok Klaster 0 (*Passive Users* - 38,67%), strategi pendekatan dititikberatkan pada penyampaian informasi secara efisien, responsif, dan tidak membebankan pengguna. Mengingat segmen ini memiliki durasi penggunaan dan tingkat interaksi yang relatif rendah, praktisi disarankan untuk mengadopsi format konten berdurasi pendek (*short-form content*) yang memikat secara visual pada 3 detik pertama serta langsung menyampaikan pesan inti (*call to action*) secara lugas tanpa menuntut partisipasi aktif seperti mengisi survei atau menulis komentar yang rumit.

Sementara itu, untuk Klaster 1 (*Active Communicators* - 42,67%) yang merupakan porsi kelompok terbesar, pendekatan difokuskan pada penguatan keterlibatan dua arah (*two-way engagement*). Karena kelompok ini secara aktif memanfaatkan media sosial sebagai sarana interaksi harian dan pembaruan informasi, format konten yang paling efektif adalah konten interaktif ringan seperti fitur *polling*, kuis cerita, pilihan ganda, hingga pertanyaan terbuka yang memicu diskusi di kolom komentar. Penggunaan narasi yang dekat dengan kehidupan sehari-hari (*relatable storytelling*) serta topik tren lokal di Kota Medan akan secara signifikan meningkatkan frekuensi keterikatan pengguna pada kelompok ini.

Adapun bagi Klaster 2 (*Heavy Users / Content Creators* - 18,66%) yang memiliki tingkat akses dan partisipasi paling intensif, strategi yang diterapkan hendaknya berorientasi pada kolaborasi dan pemberdayaan komunitas (*co-creation*). Kelompok ini tidak hanya berperan sebagai konsumen informasi, melainkan juga produsen konten aktif di ekosistem digital. Oleh karena itu, praktisi dapat melibatkan segmen ini secara strategis sebagai mitra *brand ambassador*, pengulas produk (*micro-influencer / reviewer*), maupun penggerak utama dalam kampanye berbasis *User Generated Content* (UGC) seperti tantangan tagar (*hashtag challenge*) dan kompetisi kreasi video yang mampu memperluas jangkauan pemasaran secara organik.

5. KESIMPULAN DAN SARAN

7.1 Kesimpulan

Berdasarkan hasil analisis data dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa pola penggunaan media sosial oleh Generasi Z di Kota Medan menunjukkan tingkat ketergantungan yang sangat tinggi pada platform berbasis video pendek dan visual, dengan dominasi platform TikTok mencapai 70,67% dan Instagram sebesar 26,67%, di mana mayoritas responden berada pada profil demografis mahasiswa. Melalui pendekatan *Data Mining* yang mengintegrasikan pra-pemrosesan data menggunakan *StandardScaler*, penentuan titik siku melalui *Elbow Method*, serta eksekusi algoritma *K-Means Clustering*, penelitian ini berhasil mengidentifikasi bahwa jumlah pengelompokan yang paling optimal untuk memetakan perilaku digital Generasi Z di Kota Medan adalah sebanyak tiga klaster ($k = 3$).

Pemodelan ini terbukti mampu mengelompokkan responden ke dalam tiga segmen perilaku digital yang memiliki diferensiasi karakteristik yang jelas dan terstruktur, yaitu Klaster 0 (*Passive Users*) sebesar 38,67%, Klaster 1 (*Active Communicators*) sebesar 42,67%, dan Klaster 2 (*Heavy Users*) sebesar 18,66%. Pengujian keandalan model menggunakan metrik *Silhouette Score* menghasilkan skor sebesar 0,251 yang menegaskan bahwa struktur klaster tergolong valid serta berada pada kategori *fair structure*, suatu pencapaian yang sangat memadai untuk pengolahan data perilaku manusia (*human behavioral data*) yang berciri kontinu dan dinamis. Selain itu, implementasi antarmuka berbasis web menggunakan *Streamlit* terbukti efektif dalam mengotomatisasikan seluruh tahapan analitis—mulai dari pengunggahan data kuesioner,

normalisasi variabel, visualisasi *Scatter Plot* PCA, hingga melakukan prediksi segmen perilaku untuk data individu baru secara tepat dan serbaguna.

7.2 Saran

1. Bagi Akademisi / Peneliti Selanjutnya: Disarankan menambah variabel non-numerik seperti kategori preferensi konten, serta membandingkan performa *K-Means* dengan algoritma klasterisasi lain seperti *DBSCAN* atau *Agglomerative Hierarchical Clustering*.
2. Bagi Praktisi & Pelaku Bisnis: Diharapkan memanfaatkan profil tiga segmen ini sebagai acuan dalam merancang komunikasi digital yang lebih presisi dan terarah pada target Generasi Z di Kota Medan.

DAFTAR REFERENSI

Apriyanto, A., & Sitio, A. S. H. (2025). Penerapan K-Means Clustering dalam Segmentasi Perilaku Digital Pengguna Media Sosial. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 12(1), 45–56.

Awalina, N., & Rahayu, S. (2023). Optimalisasi Strategi Pemasaran Berbasis Segmentasi Pelanggan Menggunakan Algoritma K-Means. *Jurnal Sistem Informasi Dan Bisnis Digital*, 8(2), 112–123.

Febrianty, F., Setiawan, A., & Hermawan, I. (2023). Segmentasi Perilaku Konsumen Generasi Z Menggunakan Teknik Data Mining K-Means. *Jurnal Riset Informatika*, 5(3), 201–210.

Kristanti, D., Utami, E., & Raharjo, S. (2024). Implementasi K-Means dan Silhouette Score untuk Evaluasi Segmentasi Pengguna Aplikasi Mobile. *Jurnal Sains Data Dan Informatika*, 4(1), 15–27.

Lauw, C., Handoko, H., & Wijaya, A. (2023). Integration of K-Means Clustering and Business Intelligence Dashboard for Customer Behavior Analysis. *International Journal of Information Systems and Analytics*, 11(4), 88–99.

Muhamad Akbar Fauzan, M., & Mohamad Rizan, M. (2025). Segmentasi Pengguna Aplikasi Mobile & TV Streaming Berdasarkan Pola Akses dan Engagement. *Jurnal Media Dan Teknologi Digital*, 7(1), 30–42.

Putra, R. A., Kurniawan, D., & Hidayat, T. (2025). Klasterisasi Minat Calon Mahasiswa Berdasarkan Saluran Informasi Digital Menggunakan K-Means. *Jurnal Edukasi Dan Teknologi Informasi*, 6(2), 77–86.

Riani, A. P. (2023). Analisis Penentuan Jumlah Cluster Optimal K-Means Menggunakan Metode Elbow dan Silhouette Index. *Jurnal Komputasi Dan Informasi*, 10(2), 140–149.