

Segmentation of Generation Z Spending Habits Using the K-Means Clustering Algorithm: An Empirical Study on Financial Behavior Patterns

Gunawan Sylvester^{1*}, Majid Rahardi^{2*}

* Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Indonesia

sylvestergunawan867@gmail.com¹, majid@amikom.ac.id²

Article Info

Article history:

Received 2025-10-15

Revised 2025-10-29

Accepted 2025-11-08

Keyword:

*Financial Decision-Making,
Generation Z,
K-Means Clustering,
Segmentation,
Spending Behavior.*

ABSTRACT

Generation Z, born between 1997 and 2012, exhibits unique consumption behaviors shaped by digital technology, modern lifestyles, and evolving financial decision-making patterns. This study segments their financial behavior using the K-Means clustering algorithm applied to the “Generation Z Money Spending” dataset from Kaggle. In addition to K-Means, alternative clustering algorithms—K-Medoids and Hierarchical Clustering—are evaluated to compare their effectiveness in identifying behavioral patterns. The dataset consists of 1,700 individuals with 15 numerical spending attributes, including rent, food, entertainment, education, savings, and investments. All data were normalized using Min-Max Scaling prior to clustering. The analysis identifies six distinct clusters, ranging from highly consumption-oriented groups (with higher spending on entertainment and online shopping) to financially conscious groups prioritizing savings and investments. A quantitative approach was used, incorporating exploratory data analysis, correlation testing, and the Elbow Method to determine the optimal number of clusters. The optimal cluster count of six is supported by a Davies-Bouldin Index (DBI) score of 2.412, indicating acceptable but improvable cluster separation. Each cluster displays unique characteristics: Cluster 0 (average age 20.6) focuses on savings and investments with moderate essential spending; Cluster 1 (average age 23.6) prioritizes education and higher rent expenses; Cluster 2 (average age 20.3) is digitally oriented, spending more on online shopping and entertainment; Cluster 3 (average age 25.2) demonstrates financial stability with balanced expenditures; Cluster 4 (average age 24.9) emphasizes savings and investments with moderate living costs; and Cluster 5 (average age 24.96) combines strong saving habits with balanced essential and leisure spending. Model performance was assessed using the Davies-Bouldin Index, Silhouette Score, and Calinski-Harabasz Index to ensure comprehensive evaluation of cluster quality. The findings highlight the diverse spending behaviors of Generation Z, offering valuable insights for businesses, policymakers, and financial service providers to develop targeted strategies aligned with each segment’s characteristics.



This is an open access article under the [CC-BY-SA](#) license.

I. PENDAHULUAN

Generasi Z merupakan kelompok usia muda yang lahir pada rentang tahun 1997 hingga 2012. Dalam rentang usia 18 hingga 27 tahun, individu dari generasi ini mulai aktif secara ekonomi, baik sebagai pelajar, mahasiswa, maupun pekerja pemula [1][2]. Pola konsumsi yang ditunjukkan oleh generasi

ini memiliki karakteristik unik yang sangat berbeda dibandingkan dengan generasi sebelumnya dan dipengaruhi oleh teknologi digital, media sosial, gaya hidup modern, dan kecenderungan personal dalam mengambil keputusan keuangan [3].

Pengeluaran utama yang sering dikaitkan dengan generasi Z meliputi kebutuhan pokok seperti sewa tempat tinggal,

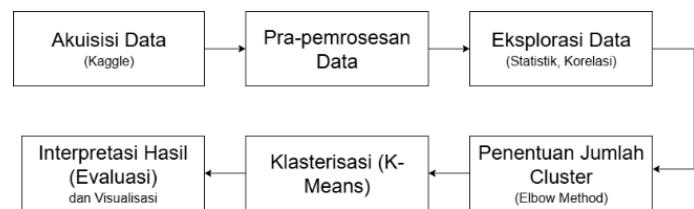
makanan, hiburan, dan pendidikan [4]. Variabel yang dipilih, seperti pengeluaran untuk sewa, makanan, hiburan, dan pendidikan, mewakili kategori utama dalam pola konsumsi Generasi Z. Kategori-kategori ini sering menjadi faktor pembeda utama dalam segmen pasar yang ditargetkan oleh bisnis keuangan. Selain itu, data ini mencakup pengeluaran untuk tabungan dan investasi, yang memberikan pandangan lengkap tentang kebiasaan keuangan mereka [5]. Kecenderungan pengeluaran ini menunjukkan variasi yang cukup besar antar individu, tergantung pada kondisi sosial, ekonomi, dan preferensi pribadi [6]. Kondisi tersebut menciptakan tantangan tersendiri bagi pelaku industri dan penyedia layanan dalam memahami serta menyusun strategi pemasaran yang efektif. Meskipun penelitian tentang perilaku konsumsi generasi ini telah banyak dilakukan, masih terdapat kekurangan dalam pemahaman terkait segmentasi perilaku belanja mereka, terutama terkait dengan prioritas finansial mereka, apakah mereka lebih fokus pada konsumsi atau pada tabungan dan investasi. Sebagian besar penelitian yang ada lebih terfokus pada aspek konsumsi dan pengeluaran untuk hiburan, dengan sedikit yang mengeksplorasi bagaimana generasi ini mengelola pengeluaran mereka untuk kebutuhan dasar seperti tabungan dan investasi, atau bagaimana perilaku keuangan mereka terbagi dalam kelompok yang berbeda. Terlebih lagi, sedikit sekali penelitian yang meneliti perilaku finansial Generasi Z di Indonesia, yang memiliki konteks sosial dan ekonomi yang berbeda dengan negara-negara Barat. Padahal, dengan memahami segmentasi ini, perusahaan dan penyedia layanan keuangan dapat merancang strategi yang lebih efektif dan sesuai dengan kebutuhan masing-masing kelompok.

Perkembangan ilmu data dan teknologi kecerdasan buatan memberikan peluang bagi peneliti untuk melakukan eksplorasi perilaku konsumen melalui pendekatan kuantitatif [7]. Salah satu metode yang dapat digunakan untuk mengidentifikasi pola pengeluaran dan membentuk segmentasi berdasarkan perilaku belanja adalah klasterisasi menggunakan algoritma K-Means [8][9]. Dengan metode ini, perilaku belanja Generasi Z dapat dianalisis dan dikelompokkan menjadi beberapa klaster berdasarkan kemiripan karakteristik [10]. Penelitian ini mengisi gap terkait pemahaman kebiasaan finansial Generasi Z dan memberikan kontribusi terhadap pengembangan strategi pemasaran yang lebih terpersonalisasi. Penelitian ini bertujuan untuk menganalisis pola belanja Generasi Z (usia 18–27 tahun) berdasarkan kategori pengeluaran seperti sewa, makanan, hiburan, dan pendidikan, serta melakukan segmentasi perilaku belanja menggunakan klasterisasi K-Means [11]. Permasalahan utama yang ingin dijawab melalui penelitian ini adalah: Bagaimana karakteristik pengeluaran dan perilaku keuangan Generasi Z dapat dikelompokkan berdasarkan atribut pengeluaran mereka? Penelitian ini memberikan kontribusi baru dengan menganalisis secara mendalam dataset "Generation Z Money Spending" dari Kaggle, yang belum banyak dieksplorasi sebelumnya, terutama dalam konteks perilaku finansial Generasi Z di Indonesia. Penelitian ini juga menggabungkan validasi hasil

klasterisasi menggunakan beberapa metrik evaluasi komprehensif seperti Davies-Bouldin Index (DBI), Silhouette Score, dan Calinski-Harabasz Index, yang lebih mendalam daripada studi sebelumnya yang hanya menggunakan Elbow Method. Selain itu, penelitian ini mengidentifikasi enam klaster perilaku finansial yang berbeda, memberikan wawasan baru bagi sektor bisnis dan keuangan digital dalam merancang produk yang lebih terpersonalisasi untuk Generasi Z, berdasarkan segmen perilaku yang teridentifikasi. Secara teoritis, penelitian ini berkontribusi pada pengembangan ilmu data mining dan analisis perilaku konsumen, sementara secara praktis, hasil segmentasi dapat membantu pelaku bisnis, penyedia layanan keuangan, dan pemasar merancang strategi pemasaran yang lebih tepat sasaran.

II. METODE

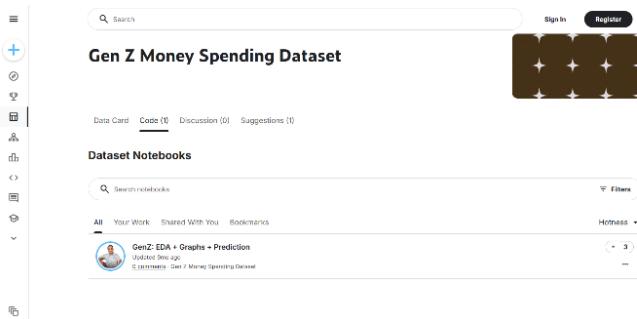
Objek penelitian dalam studi ini adalah perilaku pengeluaran keuangan Generasi Z (berusia 18–27 tahun) yang terdapat dalam dataset Generasi Z Money Spending. Dataset ini mencakup berbagai kategori pengeluaran seperti sewa, makanan, hiburan, pendidikan, dan sebagainya [12]. Adapun penelitian ini dilakukan dengan mengikuti alur sebagai berikut:



Gambar 1. Flowchart/Alur Penelitian

A. Akuisisi Data

Data yang digunakan dalam penelitian ini diambil dari platform Kaggle dalam format CSV. Kaggle adalah platform yang sering digunakan oleh peneliti dan praktisi data untuk berbagi dataset dan melakukan kompetisi data science. Dataset yang digunakan, berjudul "Gen Z Money Spending", mencakup data keuangan generasi Z, termasuk pengeluaran untuk kategori seperti sewa, makanan, hiburan, pendidikan, tabungan, dan investasi. Dataset ini memberikan gambaran yang lebih luas tentang kebiasaan konsumsi dan kebiasaan keuangan Generasi Z. Data CSV dipilih karena format ini mudah digunakan dan memiliki struktur yang terorganisir dengan baik, yang memudahkan proses analisis lebih lanjut. Dataset yang digunakan terdiri dari 1.700 baris data dengan 15 fitur, yang mencakup atribut numerik seperti pengeluaran untuk sewa, makanan, hiburan, dan kategori lainnya, serta pengeluaran untuk tabungan dan investasi. Dengan detail sebagai berikut : ID, Age, Income, Rent, Groceries, Eating Out, Entertainment, Subscription Services, Education, Online Shopping, Savings, Investments, Travel, Fitness, Miscellaneous, yang semuanya dalam bentuk USD (United States Dollar).



Gambar 2. Dataset Generasi Z Money Spending yang diambil dari Kaggle

B. Pra-pemrosesan Data

Sebelum data dianalisis, proses pra-pemrosesan data dilakukan untuk memastikan kualitas data yang lebih baik dan siap untuk analisis lebih lanjut. Dua tahap utama yang dilakukan dalam pra-pemrosesan adalah Pembersihan Data dan Standardisasi Data. Pembersihan Data termasuk menghapus kolom ID, yang mungkin tidak relevan dalam analisis klasterisasi. Kolom ID biasanya tidak memiliki nilai penting dalam menentukan pola konsumsi, jadi penghapusannya membantu menyederhanakan dataset. Selanjutnya, agar data memiliki skala yang seragam, MinMaxScaler digunakan untuk menormalkan data (Standardisasi Data). Proses ini mengubah semua nilai fitur ke dalam rentang [0, 1], yang penting dalam algoritma klasterisasi seperti K-Means [13][14]. Algoritma K-Means sangat sensitif terhadap perbedaan skala antar fitur, sehingga normalisasi ini memastikan bahwa setiap fitur memberikan kontribusi yang setara dalam perhitungan jarak antar data, meningkatkan kualitas klasterisasi, dan menghindari dominasi fitur dengan skala yang lebih besar.

C. Eksplorasi Data

Pada tahap eksplorasi data, peneliti melakukan analisis deskriptif dan korelasi antar fitur untuk memahami dataset secara lebih mendalam. Analisis deskriptif mencakup statistik dasar seperti rata-rata, median, standar deviasi, dan distribusi data untuk fitur-fitur penting seperti pengeluaran untuk sewa, makanan, hiburan, dan kategori lainnya. Hal ini membantu peneliti untuk memahami bagaimana distribusi pengeluaran di antara individu dalam dataset dan memberikan wawasan mengenai pola konsumsi mereka. Selain itu, analisis korelasi antar fitur dilakukan untuk mengetahui hubungan antara variabel pengeluaran yang satu dengan lainnya.

D. Penentuan Jumlah Cluster

Pada tahap penentuan jumlah klaster, metode Elbow Method digunakan untuk mencari nilai K (jumlah klaster) yang optimal dengan cara mengamati penurunan Within-Cluster Sum of Squares (WCSS) atau jumlah kuadrat kesalahan dalam klaster. Titik siku (elbow) pada grafik WCSS menunjukkan titik optimal untuk jumlah klaster. Namun, untuk mengevaluasi hasil klasterisasi secara lebih mendalam, peneliti juga menggunakan Davies-Bouldin Index (DBI). DBI adalah metrik yang digunakan untuk mengukur

kualitas klaster berdasarkan dua komponen utama: jarak antar centroid klaster dan sebaran dalam klaster itu sendiri [15]. Semakin kecil nilai DBI, semakin baik pemisahan antar klaster yang terbentuk, karena DBI menghitung rasio antara kedekatan dalam klaster dan jarak antar klaster [16]. Dalam penelitian ini, DBI digunakan untuk memastikan bahwa klaster yang terbentuk terpisah dengan baik dan memiliki pemisahan yang jelas antar satu dengan lainnya [17].

E. Klasterisasi

Setelah menentukan jumlah cluster optimal, algoritma K-Means diterapkan untuk mengelompokkan data berdasarkan kesamaan pengeluaran [18]. Algoritma K-Means dipilih untuk segmentasi ini karena kemampuannya dalam menangani dataset besar dengan efisien dan aplikasinya yang umum dalam pengelompokan data numerik. Meskipun K-Means dipilih dalam penelitian ini, namun ada beberapa algoritma lain yang dapat digunakan untuk segmentasi perilaku keuangan, di antaranya: 1. K-Medoids: K-Medoids adalah algoritma klasterisasi yang mirip dengan K-Means tetapi lebih stabil terhadap outlier [19]. Alih-alih menggunakan centroid untuk representasi klaster, K-Medoids menggunakan titik data aktual dalam klaster sebagai pusat klaster (medoid). Meskipun lebih robust terhadap outlier, K-Medoids memiliki waktu komputasi yang lebih tinggi dan lebih rumit dalam implementasi dibandingkan dengan K-Means, terutama untuk dataset besar seperti yang digunakan dalam penelitian ini. 2. Hierarchical Clustering: Hierarchical Clustering mengelompokkan data secara hierarkis dengan membentuk pohon klaster (dendrogram) [20]. Meskipun memberikan gambaran yang lebih rinci tentang hubungan antar data, algoritma ini memiliki keterbatasan dalam hal skalabilitas dan efisiensi komputasi pada dataset besar, karena membutuhkan perhitungan jarak antar semua pasangan data. Ini membuatnya kurang cocok untuk dataset dengan ukuran besar dan banyak fitur. 3. DBSCAN (Density-Based Spatial Clustering of Applications with Noise): DBSCAN adalah algoritma berbasis kepadatan yang dapat mengidentifikasi klaster dengan bentuk yang tidak teratur dan mengatasi masalah noise [21]. Meskipun algoritma ini dapat memberikan segmentasi yang lebih fleksibel, hasilnya sangat bergantung pada pemilihan parameter seperti epsilon (jarak maksimum antar titik dalam satu klaster) dan min_samples (jumlah titik minimum untuk membentuk klaster). DBSCAN juga cenderung tidak efisien dalam menangani dataset besar dengan dimensi yang tinggi. Meskipun algoritma lain seperti K-Medoids, Hierarchical Clustering, dan DBSCAN memiliki kelebihan tertentu, K-Means dipilih karena kesederhanaannya dalam implementasi, efisiensinya dalam waktu komputasi (terutama setelah data dinormalisasi), dan kemampuannya untuk menangani dataset besar yang digunakan dalam penelitian ini. Dengan menggunakan K-Means, peneliti dapat melakukan segmentasi secara cepat dan efisien, serta memanfaatkan berbagai metrik evaluasi seperti Elbow Method dan DBI untuk memastikan bahwa klaster yang terbentuk optimal dan terpisah dengan baik.

K-Means merupakan algoritma clustering yang membagi data ke dalam k kelompok berdasarkan jarak terdekat ke centroid masing-masing klaster [22][23]. Proses ini dilakukan secara iteratif di mana data yang lebih dekat ke centroid akan bergabung dalam satu klaster dengan tujuan meminimalkan within-cluster sum of squares (WCSS) atau variansi dalam klaster[24]. Setelah klaster terbentuk, kita dapat menganalisis karakteristik setiap klaster untuk mengidentifikasi pola yang ada dalam data [25]. Proses iteratif K-Means mencakup langkah-langkah berikut:

- Menentukan jumlah cluster (k),
- Menginisialisasi pusat cluster (centroid) secara acak,
- Mengelompokkan data ke cluster berdasarkan jarak terdekat,
- Menghitung ulang pusat cluster berdasarkan anggota baru,
- Mengulang langkah 3 dan 4 hingga konvergen.

Algoritma K-Means dihitung dengan persamaan berikut [26]:

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

dengan:

- J : fungsi objektif (within-cluster sum of squares/WCSS)
- C_i : cluster ke- i
- x : anggota data dalam cluster
- μ_i : pusat cluster ke- i

Dalam K-Means, evaluasi utama adalah posisi centroid dari masing-masing klaster. Untuk memastikan klaster yang optimal, kita menguji beberapa nilai K, mulai dari $K=2$ hingga $K=15$, dan menganalisis jarak antara centroid serta nilai DBI. Semakin kecil nilai DBI, semakin baik pemisahan antara klaster, yang menunjukkan klaster yang lebih terpisah dengan baik [27]. Evaluasi hasil klasterisasi dilakukan menggunakan beberapa metrik untuk memastikan kualitas segmentasi yang lebih baik, termasuk Davies-Bouldin Index (DBI), Silhouette Coefficient, dan Calinski-Harabasz Index. DBI mengukur kualitas pemisahan antar klaster, dengan semakin kecil nilai DBI menunjukkan pemisahan antar klaster yang lebih baik. Silhouette Coefficient menilai seberapa baik klaster-klaster tersebut terpisah dengan menghitung jarak antara data dalam klaster dan jarak antara data di klaster lain, dengan nilai yang lebih tinggi menunjukkan pemisahan klaster yang lebih baik [28]. Calinski-Harabasz Index memberikan gambaran komprehensif tentang kekompakan dan pemisahan klaster, dengan nilai yang lebih tinggi menunjukkan klaster yang lebih baik dalam hal kekompakan dalam klaster dan pemisahan antar klaster [29]. Dengan menggunakan ketiga metrik ini, penelitian ini bertujuan untuk memastikan bahwa klaster yang terbentuk memiliki kualitas yang baik dan dapat diinterpretasikan dengan jelas.

F. Interpretasi Hasil

Pada tahap Interpretasi Hasil setelah klasterisasi, peneliti menganalisis centroid dari masing-masing klaster untuk menggambarkan karakteristik pengeluaran setiap kelompok, seperti kebutuhan dasar, hiburan, tabungan, dan investasi. Proses ini membantu memahami perbedaan antara klaster-klaster yang terbentuk, misalnya ada klaster yang lebih fokus pada pengeluaran untuk sewa dan makanan, sementara klaster lain lebih mengutamakan tabungan dan investasi. Selanjutnya, peneliti mengevaluasi kualitas pemisahan antar klaster menggunakan Davies-Bouldin Index (DBI), yang mengukur seberapa baik klaster-klaster tersebut terpisah. Nilai DBI yang lebih rendah menunjukkan pemisahan yang lebih jelas antara klaster-klaster tersebut. Peneliti juga menggunakan teknik visualisasi seperti PCA (Principal Component Analysis) untuk memetakan klaster-klaster ini ke dalam ruang dua dimensi, sehingga memudahkan analisis pemisahan antar klaster. Dengan langkah-langkah ini, peneliti dapat menggali pola pengeluaran dan perilaku konsumsi dalam setiap klaster, yang memberikan wawasan berguna bagi bisnis dan lembaga keuangan dalam merancang strategi pemasaran yang lebih personal dan relevan.

III. HASIL DAN PEMBAHASAN

A. Deskripsi dan Visualisasi Data

Dataset yang digunakan dalam penelitian ini berjudul "genz_money_spends.csv" yang memuat informasi perilaku pengeluaran keuangan individu berusia 18–27 tahun, yang dikategorikan sebagai Generasi Z (Generasi Z). Data mencakup variabel-variabel seperti usia, pengeluaran untuk kebutuhan primer (sewa, makanan, pendidikan), sekunder (hiburan, langganan digital, belanja online), serta pengeluaran produktif (tabungan dan investasi).

Age	Income (USD)	U/Rent (USD)	Groceries	Eating Out	Entertainment	Subscriptions	Education	Online Shopping (USD)	Savings (USD)	Investments	Travel (USD)	Fitness (USD)	Miscellaneous (USD)
1	24	7131	544	494	303	81	21	1937	402	1587	619	886	154
2	21	5311	841	288	406	375	98	1600	262	145	2902	1032	237
3	25	4125	784	501	79	327	27	1146	590	2901	2407	1264	85
4	24	3607	402	224	259	38	60	1056	149	2157	4495	200	304
5	24	2941	424	154	475	384	15	1536	476	763	3004	1328	186
6	27	4479	1872	281	82	319	52	203	459	672	3914	1644	198
7	20	3297	469	309	278	26	44	517	75	2129	2486	163	55
8	24	5890	423	475	61	268	63	1832	546	1653	3402	553	26
9	25	5509	1358	570	167	136	71	1974	438	1721	3596	1745	75
10	22	6350	1410	174	399	87	78	516	107	2796	1677	1004	270
11	21	2894	945	376	204	53	91	442	628	1012	4039	1293	132
12	25	5781	1085	527	472	213	25	1615	675	739	1304	406	240
13	25	2769	500	232	442	344	17	1869	27	2763	4093	218	134
14	20	6522	668	442	255	74	98	688	496	2645	1593	1597	121
15	20	6861	1094	581	454	285	99	649	146	1600	3307	201	251
16	22	5214	376	158	381	61	95	632	571	2223	1361	291	70
17	19	2029	358	444	296	316	40	338	516	968	4294	1448	255
18	25	5123	730	588	78	250	19	1461	522	1129	1787	188	224
19	23	7664	937	281	210	397	65	1468	264	952	1073	146	244
20	19	4458	1823	557	284	29	85	1159	497	406	1165	196	91
21	22	4414	403	505	239	377	79	386	328	2695	3599	1255	85
22	18	6389	1307	438	166	147	29	1048	250	2474	2717	1130	24
23	27	5091	1601	420	320	198	14	505	402	1916	3153	1115	229
24	23	3647	367	163	307	212	32	1348	174	494	4126	915	239
25	26	6073	1392	426	229	111	18	1718	194	2447	3055	1784	274
26	18	7218	1468	359	422	388	67	679	134	848	659	592	273

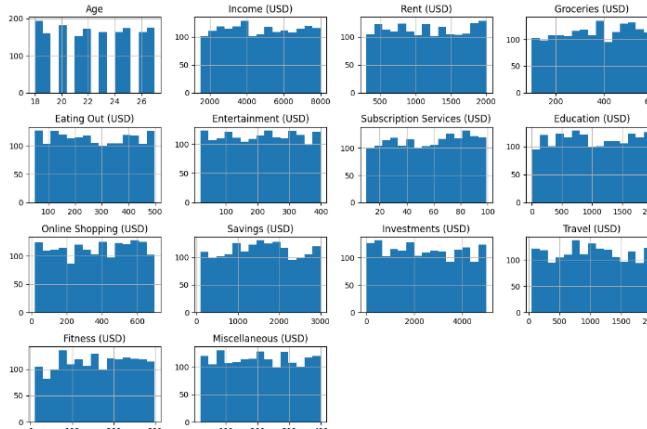
Gambar 3. Dataset "genz_money_spends.csv"

Dataset terdiri dari 1700 baris data dan 15 fitur, yang mencerminkan kategori pengeluaran individu Generasi Z dalam mata uang USD. Semua fitur selain kolom ID digunakan dalam proses analisis.

Contoh data setelah normalisasi (Min-Max):				
Age	Rent (USD)	Groceries (USD)	Eating Out (USD)	\
0	0.666667	0.143614	0.789579	0.563474
1	0.333333	0.318423	0.376754	0.792873
2	0.777778	0.284873	0.803607	0.064588
3	0.444444	0.071218	0.308617	0.552339
4	0.666667	0.072984	0.108216	0.946548
Entertainment (USD) Subscription Services (USD) Education (USD) \				
0	0.160950	0.123596	0.970441	\
1	0.936675	0.988764	0.831663	
2	0.810026	0.191011	0.574148	
3	0.047493	0.876404	0.529058	
4	0.960422	0.056180	0.769539	
Online Shopping (USD) Savings (USD) Investments (USD) Travel (USD) \				
0	0.562592	0.512935	0.123825	0.428278
1	0.356486	0.015523	0.580516	0.583342
2	0.833579	0.997240	0.481496	0.632905
3	0.206186	0.709555	0.890578	0.080720
4	0.671576	0.228700	0.600920	0.655527
Fitness (USD) Miscellaneous (USD)				
0	0.498270	0.715040		
1	0.785467	0.728232		
2	0.259516	0.361478		
3	0.768166	0.952507		
4	0.608997	0.345646		

Gambar 4. Tampilan Data Setelah Dilakukan Normalisasi

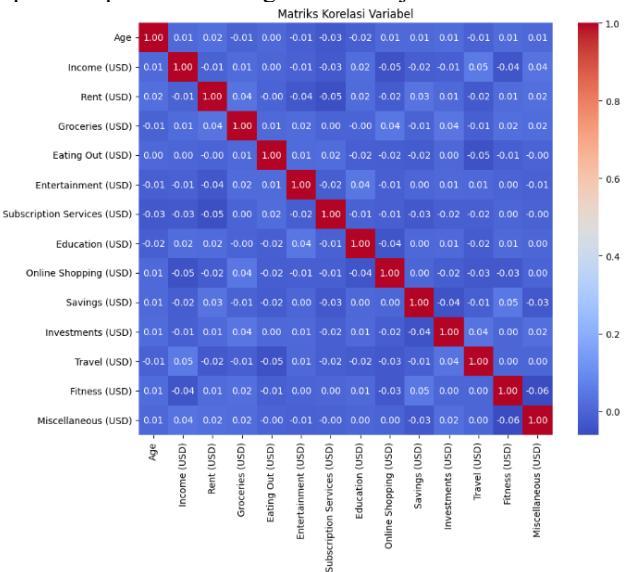
Setelah normalisasi data, data di visualkan dengan visualisasi sebagai berikut :



Gambar 5. Visualisasi Histogram Data

Visualisasi histogram menunjukkan beberapa temuan penting terkait pola pengeluaran Generasi Z. Distribusi usia yang cukup merata antara 18 hingga 26 tahun mencerminkan fase transisi Generasi Z dari pendidikan ke dunia kerja. Hal ini menunjukkan bahwa pengeluaran dalam kategori seperti sewa dan makanan dipengaruhi oleh status sosial ekonomi yang lebih variatif di antara usia-usia tersebut, yang sejalan dengan temuan dalam literatur bahwa pengeluaran untuk kebutuhan dasar meningkat seiring dengan meningkatnya status pekerjaan dan pendidikan. Pendapatan mereka tersebar antara 1.500 hingga 8.000 USD, tanpa dominasi nilai tertentu, yang mengindikasikan variasi ekonomi yang luas. Pengeluaran untuk kebutuhan pokok, seperti sewa (Rent) dan bahan makanan (Groceries), terdistribusi merata, meskipun terdapat beberapa lonjakan pada nilai tertentu, mencerminkan variasi biaya hidup yang dipengaruhi oleh tempat tinggal dan gaya hidup. Pengeluaran untuk makan di luar (Eating Out) dan hiburan (Entertainment) juga terdistribusi merata, namun menunjukkan bahwa konsumsi di kedua kategori ini adalah pengeluaran yang konsisten di semua kelompok. Sebagian

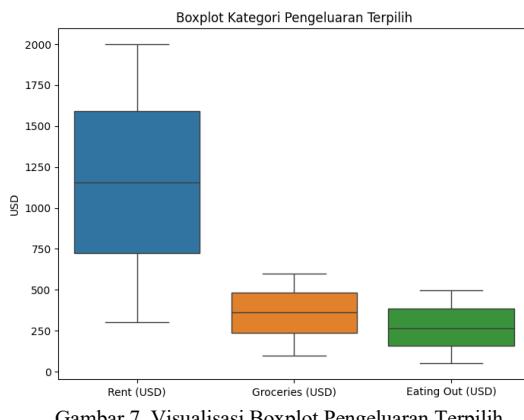
besar pengeluaran untuk layanan digital berada di rentang 20-80 USD, menandakan banyaknya Generasi Z yang menganggarkan secara konsisten untuk langganan digital. Terkait pengeluaran edukasi, terlihat variasi yang tinggi, di mana sebagian besar responden membayar mahal untuk pendidikan, sementara yang lainnya tidak mengeluarkan biaya pendidikan sama sekali, yang bisa dipengaruhi oleh status pelajar atau pekerjaan. Pengeluaran untuk kebugaran (Fitness) tidak merata, menunjukkan bahwa sebagian besar Generasi Z tidak mengeluarkan banyak untuk kebugaran. Untuk gaya hidup konsumtif, belanja online (Online Shopping) menunjukkan distribusi tinggi di semua rentang, mengindikasikan bahwa pengeluaran untuk belanja online cukup umum. Beberapa responden menghabiskan banyak uang untuk perjalanan (Travel), hingga 2000 USD, meskipun sebagian besar tetap berada di bawah 1000 USD. Pengeluaran untuk Miscellaneous terlihat merata di banyak rentang, menunjukkan adanya pengeluaran tak terduga atau kebutuhan kecil lainnya. Di sisi lain, pengeluaran untuk tabungan (Savings) terdistribusi rata, yang menunjukkan kesadaran menabung yang cukup tinggi meskipun jumlahnya bervariasi. Begitu juga dengan investasi (Investments), yang terdistribusi cukup lebar, mengindikasikan adanya kelompok Generasi Z yang sudah aktif berinvestasi. Kesimpulannya, Generasi Z memiliki gaya hidup konsumtif, terutama untuk makan di luar, belanja online, hiburan, dan langganan digital. Terdapat variasi besar dalam pendapatan, tabungan, dan investasi, yang menunjukkan segmentasi yang potensial untuk klasterisasi dan prediksi perilaku keuangan lebih lanjut.



Gambar 6. Visualisasi Heatmap Pengeluaran Gen-Z

Heatmap ini menggambarkan hubungan linier antar variabel dengan nilai berkisar antara -1 hingga 1. Nilai 1.00 menunjukkan korelasi sangat kuat positif, 0.00 menunjukkan tidak ada korelasi linier, dan -1.00 menunjukkan korelasi sangat kuat negatif. Warna merah menunjukkan korelasi kuat, sedangkan biru tua menunjukkan korelasi lemah atau tidak ada korelasi. Hasil analisis menunjukkan bahwa semua nilai diagonal bernilai 1.00 karena setiap variabel berkorelasi

sempurna dengan dirinya sendiri, namun korelasi antar variabel lainnya sangat rendah, sebagian besar berada di rentang -0.05 hingga 0.05, yang berarti tidak ada hubungan linier yang signifikan antar pengeluaran dan variabel lainnya. Sebagai contoh, tabungan hanya berkorelasi 0.01–0.05 dengan variabel lain, menunjukkan bahwa tabungan tidak dipengaruhi langsung oleh satu jenis pengeluaran spesifik. Pendapatan juga berkorelasi negatif sangat lemah dengan pengeluaran untuk sewa, pendidikan, hiburan, dan lainnya, dengan nilai korelasi sekitar -0.01 hingga -0.05. Implikasinya adalah tidak ada variabel pengeluaran yang saling berkaitan erat secara linier dalam pola belanja Generasi Z, yang membuat model prediksi berbasis regresi linier mungkin kurang akurat. Oleh karena itu, analisis lanjutan yang lebih cocok adalah klasterisasi, seperti K-Means, yang fokus pada pengelompokan perilaku belanja, bukan pada hubungan antar variabel. Visualisasi histogram dan heatmap juga menunjukkan variasi pengeluaran dan ketidakterkaitan antar pengeluaran satu sama lain, yang memperkuat kebutuhan untuk menggunakan metode klasterisasi.



Gambar 7. Visualisasi Boxplot Pengeluaran Terpilih

Pada kategori Rent (USD), median pengeluaran sewa bulanan sekitar 1.150 USD, dengan rentang antar kuartil (IQR-Interquartile Range) yang cukup lebar, yaitu sekitar 750 – 1.600 USD, menunjukkan variasi biaya sewa yang cukup tinggi. Nilai minimum tercatat sekitar 300 USD, sementara maksimum mencapai 2.000 USD, yang menunjukkan adanya individu dengan pengeluaran sewa sangat rendah dan sangat tinggi. Kesimpulannya, sewa merupakan kategori pengeluaran dengan variabilitas terbesar dan menyerap porsi pengeluaran paling tinggi. Untuk Groceries (USD), median pengeluaran tercatat sekitar 350 USD, dengan IQR antara 250 – 500 USD, yang menunjukkan variasi yang lebih kecil dibandingkan sewa. Rentang nilai pengeluaran berkisar antara 100 – 600 USD. Kesimpulannya, pengeluaran untuk kebutuhan pokok relatif lebih stabil dan tidak terlalu bervariasi antar individu.

Pada kategori Eating Out (USD), median pengeluaran berada pada angka 250 USD, dengan IQR sekitar 150 – 400 USD dan rentang nilai antara 50 – 500 USD. Meskipun ada variasi, pengeluaran makan di luar lebih fleksibel dan bisa disesuaikan dengan gaya hidup individu, tetapi variasinya

lebih kecil dibandingkan dengan sewa. Pengeluaran untuk sewa mendominasi pengeluaran bulanan dan menunjukkan perbedaan signifikan antar individu, dengan variasi yang sangat tinggi. Sebaliknya, pengeluaran untuk groceries dan makan di luar lebih kecil dan memiliki rentang yang lebih sempit, yang menunjukkan konsistensi antar individu dalam pengeluaran ini. Pola ini menunjukkan bahwa biaya hidup paling dipengaruhi oleh pengeluaran sewa, sementara kebutuhan sehari-hari seperti belanja bahan makanan dan makan di luar cenderung lebih terkendali.

B. Hasil Klasterisasi K-Means

Proses klasterisasi menggunakan metode K-Means menghasilkan 6 klaster terbaik berdasarkan analisis visual dan metrik evaluasi seperti SSE (Sum of Squared Errors) dan Silhouette Score yang telah di evaluasi dengan DBI (Davies-Bouldin Index).

TABEL I
CONTOH BEBERAPA BARIS DENGAN KLASTER

Age	Income (USD)	Cluster	Makna
24	7131	4	Individu usia 24 tahun dengan penghasilan tinggi (USD 7131) masuk ke Cluster 4, yang kemungkinan mewakili kelompok dengan penghasilan menengah-tinggi.
21	5311	0	Individu usia 21 tahun dengan penghasilan cukup tinggi (USD 5311) masuk ke Cluster 0, yang tampaknya menampung usia muda dengan variasi penghasilan yang luas.
25	4125	4	Individu usia 25 tahun dengan penghasilan sedang (USD 4125) juga masuk Cluster 4, menunjukkan bahwa klaster ini menampung usia lebih matang dengan penghasilan menengah-atas.
22	3882	5	Individu usia 22 tahun dengan penghasilan sedang (USD 3882) masuk ke Cluster 5, kemungkinan kelompok usia muda dengan penghasilan menengah.
24	2941	1	Individu usia 24 tahun dengan penghasilan lebih rendah (USD 2941) masuk ke Cluster 1, yang mungkin mewakili usia dewasa muda dengan penghasilan relatif rendah.

Centroid setiap klaster (skala asli):					
	Age	Rent (USD)	Groceries (USD)	Eating Out (USD)	Entertainment (USD)
0	20.61	895.46	427.96	229.05	255.26
1	23.65	1295.91	325.83	221.70	247.85
2	20.34	1171.33	292.26	270.09	179.52
3	20.10	1321.70	379.59	313.02	198.63
4	24.97	1044.59	363.99	274.38	212.02
5	24.96	1177.49	365.76	308.08	178.61
Subscription Services (USD)					
	Age	Rent (USD)	Groceries (USD)	Online Shopping (USD)	Entertainment (USD)
0		71.23	948.93	462.30	
1		35.35	1419.07	297.37	
2		71.03	1311.73	273.76	
3		38.38	671.27	396.06	
4		40.16	930.24	423.62	
5		76.88	810.65	348.13	
Savings (USD)					
	Age	Rent (USD)	Groceries (USD)	Travel (USD)	Fitness (USD)
0		1590.70	2867.11	1306.33	205.67
1		1509.18	3417.01	1047.15	187.54
2		1627.74	1629.23	841.80	143.73
3		1587.73	2531.73	1039.49	116.69
4		1874.02	1497.74	1000.19	162.27
5		1164.34	3013.19	939.43	148.76
Miscellaneous (USD)					
	Age	Rent (USD)	Groceries (USD)	Travel (USD)	Entertainment (USD)
0		170.14			
1		267.00			
2		188.16			
3		246.10			
4		155.64			
5		227.75			

Gambar 8. Hasil Klasterisasi

Berdasarkan hasil klasterisasi, klaster dengan usia lebih muda cenderung lebih mengutamakan pengeluaran untuk hiburan dan belanja online, yang mencerminkan gaya hidup konsumtif yang dipengaruhi oleh adopsi teknologi dan media sosial. Sementara itu, klaster yang lebih tua lebih fokus pada pengeluaran untuk kebutuhan dasar, tabungan, dan investasi, yang menunjukkan kecenderungan yang lebih kuat terhadap literasi keuangan dan perencanaan keuangan jangka panjang. Temuan ini menunjukkan adanya keterkaitan langsung antara pengaruh sosial dan gaya hidup digital yang berkembang pesat di kalangan Generasi Z. Klaster dengan usia lebih muda (seperti Cluster 0 dan Cluster 2) lebih banyak menghabiskan uang untuk hiburan dan belanja online, yang dapat dijelaskan melalui Teori Adopsi Teknologi yang mengemukakan bahwa individu muda cenderung lebih cepat mengadopsi teknologi baru, seperti e-commerce dan platform hiburan digital. Hal ini juga dipengaruhi oleh fenomena Fear of Missing Out (FOMO) yang sangat kuat pada generasi ini, mendorong mereka untuk mengikuti tren sosial dan konsumsi berbasis media sosial. Sebaliknya, klaster yang lebih tua (seperti Cluster 3 dan Cluster 5) lebih fokus pada pengeluaran untuk kebutuhan dasar dan investasi, yang mengindikasikan adanya perubahan prioritas finansial yang muncul seiring bertambahnya usia dan pengalaman hidup. Mereka lebih cenderung mengikuti Teori Siklus Hidup Konsumsi yang menekankan pentingnya menabung dan berinvestasi seiring bertambahnya usia dan stabilitas finansial. Temuan ini juga konsisten dengan teori Theory of Planned Behavior, yang menunjukkan bahwa keputusan finansial sering kali dipengaruhi oleh norma sosial dan kontrol perilaku yang dipersepsi. Klaster dengan usia lebih muda (misalnya, Cluster 2) menunjukkan kecenderungan untuk lebih banyak berbelanja online dan menghabiskan uang untuk hiburan. Hal ini mungkin disebabkan oleh adopsi teknologi yang tinggi di kalangan generasi muda, yang memudahkan mereka untuk terlibat dalam konsumsi digital. Sementara itu, klaster yang lebih tua (Cluster 3 dan Cluster 4) lebih menitikberatkan pada tabungan dan investasi, yang dapat dijelaskan melalui life-cycle hypothesis, yang menyatakan bahwa semakin tua usia

individu, semakin besar kemungkinan mereka untuk menabung dan berinvestasi sebagai bagian dari perencanaan finansial untuk masa depan.

1) Centroid Setiap Klaster

Centroid menunjukkan nilai rata-rata pengeluaran untuk setiap kategori yang digunakan dalam klasterisasi untuk setiap klaster. Berdasarkan hasil centroid, kita dapat melihat karakteristik umum dari masing-masing klaster.

Cluster 0: Individu dalam Cluster 0 memiliki usia rata-rata 20.61 tahun dan cenderung mengutamakan kebutuhan dasar seperti tempat tinggal dan bahan makanan, dengan pengeluaran untuk sewa (895.46 USD) dan bahan makanan (427.96 USD). Meskipun demikian, mereka juga mengalokasikan dana untuk hiburan dan makan di luar, dengan pengeluaran moderat untuk kategori Entertainment (255.26 USD) dan Eating Out (229.05 USD). Individu dalam klaster ini juga terlihat fokus pada tabungan dan investasi, dengan pengeluaran untuk Savings (1590.70 USD) dan Investments (2867.11 USD), menunjukkan kecenderungan mereka untuk menyisihkan lebih banyak uang untuk masa depan finansial mereka.

Cluster 1: Dengan usia rata-rata 23.65 tahun, individu dalam Cluster 1 cenderung tinggal di tempat dengan biaya hidup yang lebih tinggi, terbukti dengan pengeluaran mereka yang lebih tinggi untuk sewa (1295.91 USD) dibandingkan Cluster 0. Meskipun pengeluaran untuk bahan makanan (325.83 USD) sedikit lebih rendah, pengeluaran untuk hiburan dan makan di luar (Eating Out: 221.70 USD dan Entertainment: 247.85 USD) masih moderat. Mereka cukup berorientasi pada tabungan dan investasi, dengan pengeluaran untuk Savings (1509.18 USD) dan Investments (3417.01 USD), yang mencerminkan fokus pada masa depan finansial mereka.

Cluster 2: Individu dalam Cluster 2 memiliki usia rata-rata 20.34 tahun, mirip dengan Cluster 0, dan pengeluaran mereka untuk sewa (1171.33 USD) serta bahan makanan (292.26 USD) lebih moderat dibandingkan Cluster 0. Mereka lebih cenderung menghabiskan uang untuk makan di luar (270.09 USD) dan hiburan (179.52 USD), yang menunjukkan bahwa mereka memiliki gaya hidup konsumtif. Meskipun demikian, mereka juga menyisihkan uang untuk tabungan dan investasi, dengan pengeluaran untuk Savings (1627.74 USD) dan Investments (1629.23 USD), menunjukkan perhatian terhadap keuangan pribadi. Cluster 3: Cluster 3 terdiri dari individu dengan usia rata-rata 20.10 tahun dan pengeluaran untuk sewa (1321.70 USD) yang cukup tinggi, serta pengeluaran untuk bahan makanan (379.59 USD) yang lebih besar dibandingkan Cluster 0 dan Cluster 2. Pengeluaran mereka untuk makan di luar (313.02 USD) dan hiburan (198.63 USD) cukup tinggi, mencerminkan gaya hidup yang lebih konsumtif. Meskipun demikian, mereka tetap fokus pada tabungan dan investasi, dengan pengeluaran untuk Savings (1587.73 USD) dan Investments (2531.73 USD), yang menunjukkan keseimbangan antara pengeluaran gaya hidup dan perencanaan keuangan. Cluster 4: Individu dalam Cluster 4 berusia rata-rata 24.97 tahun dan cenderung memiliki pengeluaran untuk sewa (1044.59 USD) yang lebih

rendah dibandingkan Cluster 1 dan Cluster 3, meskipun masih cukup signifikan. Pengeluaran mereka untuk bahan makanan (363.99 USD) cukup moderat, dan mereka menghabiskan uang untuk makan di luar (274.38 USD) dan hiburan (212.02 USD), menunjukkan gaya hidup yang aktif dalam hiburan. Fokus mereka lebih besar pada tabungan dan investasi, dengan pengeluaran untuk Savings (1874.02 USD) dan Investments (1497.74 USD), yang menunjukkan prioritas terhadap keuangan jangka panjang dibandingkan pengeluaran konsumtif. Cluster 5: Individu dalam Cluster 5 memiliki usia rata-rata 24.96 tahun, mirip dengan Cluster 4, dan pengeluaran mereka untuk sewa (1177.49 USD) dan bahan makanan (365.76 USD) cukup moderat. Namun, pengeluaran untuk makan di luar (308.08 USD) lebih tinggi dibandingkan dengan klaster lainnya. Pengeluaran mereka untuk hiburan (178.61 USD) sedikit lebih rendah, menunjukkan mereka lebih fokus pada pengeluaran yang lebih produktif. Mereka juga menunjukkan kecenderungan kuat untuk berfokus pada investasi, dengan pengeluaran untuk Investments (3013.19 USD), meskipun tabungan mereka (1164.34 USD) relatif lebih rendah dibandingkan dengan klaster lainnya.

2) *Ukuran Setiap Klaster*

TABEL II
UKURAN SETIAP KLASTER

Cluster	Jumlah Anggota
Cluster 0	242
Cluster 1	251
Cluster 2	316
Cluster 3	293
Cluster 4	293
Cluster 5	305

Ukuran setiap Klaster menunjukkan bahwa Cluster 2 memiliki jumlah anggota terbesar (316), diikuti oleh Cluster 5 (305 anggota). Hal ini mengindikasikan bahwa kedua klaster ini mungkin lebih representatif dari populasi secara keseluruhan. Sementara itu, Cluster 0 memiliki jumlah anggota terkecil (242), yang menunjukkan bahwa klaster ini lebih terfokus pada segmen tertentu dengan pengeluaran yang berbeda dibandingkan dengan kelompok lainnya. Kesimpulannya, Perbedaan pengeluaran pada Generasi Z menunjukkan bahwa klaster dengan usia lebih muda (Cluster 0, 2) cenderung menghabiskan lebih banyak untuk hiburan dan makan di luar, meskipun masih fokus pada tabungan dan investasi. Sementara itu, klaster dengan usia lebih tua (Cluster 1, 4, 5) memiliki pengeluaran yang lebih seimbang antara kebutuhan dasar, investasi, dan tabungan. Semua klaster menunjukkan pengeluaran signifikan untuk tabungan dan investasi, namun klaster yang lebih tua (Cluster 1, 4, 5) lebih memfokuskan dana mereka untuk investasi. Di sisi lain, klaster dengan pengeluaran tinggi untuk makan di luar dan hiburan (misalnya Cluster 2 dan Cluster 3) menggambarkan individu dengan gaya hidup yang lebih konsumtif, cenderung menghabiskan uang lebih banyak pada hiburan.

3) *Perhitungan selisih (Euclidean Distance) antara Centroid dan menampilkan DBI untuk setiap K*

```
Σ selisih (Euclidean Distance) antara centroid:
[[0.          0.76759334 0.65299552 0.67548055 0.73483626 0.71413495]
 [0.76759334 0.          0.74944222 0.69587591 0.65343008 0.68551173]
 [0.65299552 0.74944222 0.          0.62930246 0.72793772 0.69726556]
 [0.67548055 0.69587591 0.62930246 0.          0.69479003 0.74172641]
 [0.73483626 0.65343008 0.72793772 0.69479003 0.          0.63216476]
 [0.71413495 0.68551173 0.69726556 0.74172641 0.63216476 0.]]
```

Gambar 9. Selisih (Euclidean Distance) Antar Centroid

Berdasarkan matriks selisih Euclidean antara centroid klaster yang diberikan, kita dapat menganalisis perbedaan antara klaster-klaster berdasarkan jarak Euclidean. Jarak Euclidean antar centroid menunjukkan seberapa besar perbedaan karakteristik pengeluaran antar klaster. Semakin kecil jarak Euclidean antar centroid, semakin mirip dua klaster dalam hal karakteristik pengeluarannya.

TABEL III
MATRIKS SELISIH (EUCLIDEAN DISTANCE) ANTAR CENTROID

Cluster	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster 0	0.000	0.768	0.653	0.675	0.735	0.714
Cluster 1	0.768	0.000	0.749	0.696	0.653	0.686
Cluster 2	0.653	0.749	0.000	0.629	0.728	0.697
Cluster 3	0.675	0.696	0.629	0.000	0.695	0.742
Cluster 4	0.735	0.653	0.728	0.695	0.000	0.632
Cluster 5	0.714	0.686	0.697	0.742	0.632	0.000

Analisis jarak Euclidean menunjukkan perbedaan signifikan antara klaster-klaster berdasarkan pola pengeluaran mereka. Misalnya, Jarak Euclidean antara Cluster 0 dan Cluster 1 yang mencapai 0.768 menunjukkan bahwa perbedaan dalam pola pengeluaran sangat dipengaruhi oleh kebutuhan dasar dan gaya hidup yang berbeda. Cluster 0 yang lebih muda dengan pengeluaran tinggi untuk hiburan memiliki karakteristik yang sangat berbeda dengan Cluster 1, yang menunjukkan pengeluaran lebih tinggi untuk kebutuhan dasar seperti sewa. Pengeluaran untuk hiburan pada Cluster 1 yang lebih rendah ini menandakan adanya prioritas finansial yang berbeda, mungkin didorong oleh kebutuhan keluarga atau perencanaan masa depan yang lebih matang. Sementara itu, jarak antara Cluster 1 dan Cluster 2 adalah 0.749, yang menunjukkan perbedaan yang masih signifikan, tetapi tidak sekutu perbedaan antara Cluster 0 dan Cluster 1. Perbedaan ini mungkin mencerminkan pola konsumsi yang berbeda dalam kategori seperti Eating Out, Groceries, dan Entertainment. Jarak antara Cluster 2 dan Cluster 3 adalah 0.629, yang menunjukkan perbedaan lebih kecil, artinya kedua klaster ini lebih mirip satu sama lain dalam pola pengeluaran, terutama pada kategori Eating Out dan Groceries. Jarak antara Cluster 3 dan Cluster 4 adalah 0.695, yang menunjukkan perbedaan, namun keduanya mungkin memiliki pola pengeluaran yang cukup serupa dalam kategori

tertentu, meskipun ada perbedaan dalam pengeluaran untuk Rent atau Investments. Jarak antara Cluster 4 dan Cluster 5 adalah 0.632, yang menunjukkan keduanya cukup mirip meskipun ada beberapa perbedaan, yang kemungkinan terkait dengan gaya hidup atau tabungan. Terakhir, jarak antara Cluster 0 dan Cluster 5 adalah 0.714, yang menunjukkan perbedaan moderat, yang berarti keduanya memiliki pola pengeluaran yang berbeda dalam beberapa kategori, meskipun ada beberapa kesamaan di kategori lain. Kesimpulannya, klaster dengan jarak Euclidean lebih dekat, seperti Cluster 2 vs Cluster 3 dan Cluster 4 vs Cluster 5, memiliki kesamaan lebih besar dalam pola pengeluaran, terutama pada kategori seperti Groceries atau Eating Out. Sebaliknya, Cluster 0 vs Cluster 1 menunjukkan jarak yang paling jauh (0.768), mencerminkan perbedaan besar dalam pola pengeluaran mereka, yang mungkin mencerminkan perbedaan dalam gaya hidup atau pengeluaran kebutuhan dasar.

4) Menampilkan nilai Davies-Bouldin Index (DBI) untuk setiap K

Nilai DBI pada tiap K (2 hingga 15):	
K = 2,	DBI = 3.555
K = 3,	DBI = 3.434
K = 4,	DBI = 3.356
K = 5,	DBI = 3.140
K = 6,	DBI = 2.958
K = 7,	DBI = 2.837
K = 8,	DBI = 2.719
K = 9,	DBI = 2.654
K = 10,	DBI = 2.566
K = 11,	DBI = 2.569
K = 12,	DBI = 2.482
K = 13,	DBI = 2.412
K = 14,	DBI = 2.438
K = 15,	DBI = 2.412

Gambar 10. Nilai DBI Pada Setiap K

Nilai Davies-Bouldin Index (DBI) memberikan informasi mengenai kualitas klaster yang terbentuk, berdasarkan jarak antar centroid dan dispersion (sebaran) dalam setiap klaster. DBI adalah metrik untuk mengevaluasi kualitas klasterisasi, di mana semakin kecil nilai DBI, semakin baik kualitas klaster tersebut. Nilai DBI yang lebih kecil menunjukkan bahwa klaster lebih padat dan terpisah dengan baik. DBI dihitung berdasarkan dua komponen utama: jarak antar centroid klaster, di mana semakin jauh centroid antar klaster, semakin baik pemisahan antar klaster, dan dispersion dalam klaster, yang semakin kecil penyebaran klaster, semakin baik kualitas klaster tersebut. Berdasarkan nilai DBI untuk jumlah klaster k=2 hingga k=15, dapat disimpulkan bahwa peningkatan jumlah klaster mengurangi DBI, dengan nilai DBI yang menurun dari 3.555 pada k=2 menjadi 2.412 pada k=15. Ini menunjukkan bahwa dengan semakin banyak klaster, pemisahan antar klaster menjadi lebih jelas dan klaster lebih homogen. Penurunan DBI yang signifikan terjadi antara k=2 dan k=6, di mana nilai DBI turun dari 3.555 ke 2.958, menunjukkan bahwa menambah jumlah klaster pada titik ini memberikan perbaikan dalam kualitas pemisahan. Setelah

k=6, penurunan DBI semakin kecil, menunjukkan bahwa penambahan jumlah klaster lebih lanjut memberikan manfaat yang semakin terbatas. Setelah k=10, nilai DBI stagnan, menunjukkan bahwa penambahan klaster lebih lanjut tidak memberikan peningkatan signifikan dalam kualitas pemisahan. Jadi, evaluasi hasil klasterisasi menggunakan Davies-Bouldin Index (DBI) menunjukkan pemisahan antar klaster yang kurang optimal. Dengan DBI sebesar 2.412, pemisahan antar klaster dalam penelitian ini masih terbilang lemah, yang menunjukkan bahwa klaster-klaster tersebut belum terpisah secara optimal dengan adanya potensi tumpang tindih yang signifikan antar klaster. Ini menunjukkan bahwa meskipun klasterisasi berhasil mengelompokkan data, pemisahan antar klaster masih bisa ditingkatkan untuk mendapatkan segmentasi yang lebih jelas dan terpisah. Oleh karena itu, meskipun klasterisasi sudah memberikan wawasan yang berguna, perbaikan lebih lanjut pada pemisahan klaster mungkin diperlukan untuk meningkatkan kualitas segmentasi. Kesimpulannya, jumlah klaster optimal berdasarkan DBI berada antara k=6 hingga k=10, karena pada titik ini, DBI cukup rendah dan stabil, yang menunjukkan klaster-klaster sudah terpisah dengan baik dan kualitasnya maksimal. Peningkatan lebih lanjut dalam jumlah klaster, misalnya hingga k=15, tidak memberikan banyak perbaikan, yang menunjukkan bahwa k=10 atau k=6 sudah memadai. Dari hasil penelitian yang diperoleh, kita bisa memilih k = 6 atau k = 10 sebagai jumlah klaster optimal, karena DBI pada kedua nilai ini cukup rendah dan stabil.

5) Informasi centroid dan klaster Centroid Setiap Klaster (Skala Asli)

Centroid setiap klaster (skala asli):				
	Age	Rent	Groceries	Eating Out
0	26.64	711.08	324.68	363.62
1	19.93	761.56	291.04	229.56
2	19.95	1213.30	282.75	214.57
3	24.14	1568.47	449.18	324.65
4	24.10	1261.73	427.29	271.52
5	21.29	890.07	388.63	150.99
6	21.73	823.49	435.15	331.54
7	19.58	1283.01	272.27	495.88
8	21.13	1618.56	428.58	271.89
9	25.40	1056.68	252.97	166.92
10	23.50	901.45	431.99	318.45
11	25.07	953.54	288.73	332.37
12	22.98	1603.83	320.70	174.04
13	25.58	1335.81	282.86	355.05
14	21.29	1341.04	495.50	222.51

	Subscription Services	Education	Online Shopping	Entertainment
0	64.45	633.87	510.56	
1	38.65	1231.44	237.15	
2	76.27	1371.32	439.32	
3	72.77	958.35	273.25	
4	65.61	745.98	226.34	
5	60.03	740.38	303.14	
6	33.20	753.08	482.40	
7	60.25	968.33	234.64	
8	34.73	1144.07	450.02	
9	48.19	961.74	493.02	
10	71.78	1251.22	451.36	
11	49.56	1335.91	242.47	
12	36.79	915.68	281.29	
13	56.44	982.97	317.32	
14	73.37	965.73	472.32	

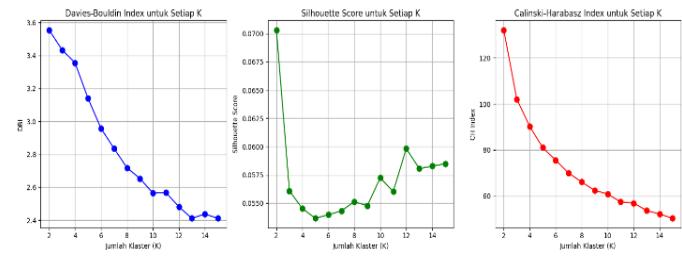
Gambar 11. Informasi centroid dan klaster Centroid Setiap Klaster (Skala Asli)

Dari data centroid yang diberikan, kita dapat melihat pengeluaran rata-rata di setiap kategori untuk setiap klaster. Setiap klaster menggambarkan kelompok dengan pola pengeluaran yang serupa, berdasarkan kategori pengeluaran yang relevan. Berikut adalah analisis yang lebih mendalam: Cluster 0 terdiri dari individu dengan usia rata-rata 20.64

tahun, yang memiliki gaya hidup aktif dengan pengeluaran yang lebih tinggi untuk hiburan dan makan di luar, dengan pengeluaran untuk Eating Out (303.62 USD) dan Entertainment (240.07 USD). Meskipun demikian, mereka tetap menyisihkan dana yang cukup besar untuk tabungan (1590.70 USD) dan investasi (3471.42 USD), menunjukkan bahwa mereka tetap berfokus pada perencanaan keuangan. Cluster 1, dengan usia rata-rata 19.93 tahun, memiliki pengeluaran yang sedikit lebih tinggi untuk sewa (761.56 USD) dibandingkan Cluster 0, tetapi pengeluaran untuk kebutuhan lain seperti Eating Out (229.56 USD) dan Entertainment (302.97 USD) sedikit lebih rendah. Klaster ini juga menunjukkan komitmen pada tabungan (1507.95 USD) dan investasi (3062.51 USD), meskipun pengeluaran hiburan mereka lebih tinggi dibandingkan dengan Cluster 0. Cluster 2, dengan usia rata-rata 19.95 tahun, memiliki pengeluaran lebih tinggi untuk sewa (1213.30 USD) tetapi lebih rendah untuk Eating Out (214.57 USD) dan Entertainment (205.72 USD). Meskipun tabungan (1427.91 USD) dan investasi (1343.90 USD) lebih rendah, klaster ini menunjukkan bahwa mereka lebih fokus pada kebutuhan dasar. Cluster 3, dengan usia rata-rata 24.14 tahun, menggambarkan individu yang lebih tua dengan pengeluaran lebih tinggi untuk sewa (1568.47 USD) dan bahan makanan (449.18 USD), namun dengan pengeluaran hiburan (161.69 USD) yang lebih rendah. Mereka tetap memprioritaskan tabungan (2295.52 USD) dan investasi (1832.26 USD), menunjukkan keseimbangan antara kebutuhan dasar dan perencanaan keuangan jangka panjang. Cluster 4, dengan usia rata-rata 24.10 tahun, memiliki pengeluaran yang lebih tinggi untuk kebutuhan dasar seperti sewa (1261.73 USD) dan bahan makanan (427.29 USD), namun lebih moderat dalam hal hiburan (Entertainment: 146.04 USD). Mereka lebih fokus pada tabungan (1005.87 USD) dan investasi (3507.02 USD), dengan perhatian lebih pada perencanaan keuangan jangka panjang. Cluster 5, dengan usia rata-rata 21.29 tahun, memiliki pengeluaran lebih rendah untuk sewa (890.07 USD) tetapi lebih tinggi untuk bahan makanan (388.63 USD). Pengeluaran mereka untuk hiburan (Eating Out: 150.99 USD, Entertainment: 98.26 USD) lebih rendah dibandingkan klaster lainnya, dan mereka juga menunjukkan fokus yang kuat pada tabungan (1413.61 USD) dan investasi (2255.53 USD).

Kesimpulannya, klaster dengan usia lebih muda (Cluster 0, 1, 2) cenderung memiliki pengeluaran lebih tinggi untuk hiburan dan makan di luar, tetapi mereka tetap menyisihkan uang untuk tabungan dan investasi. Klaster dengan usia lebih tua (Cluster 3, 4, 5) lebih fokus pada pengeluaran untuk kebutuhan dasar seperti tempat tinggal dan bahan makanan, namun tetap berfokus pada tabungan dan investasi. Semua klaster menunjukkan tingkat pengeluaran yang cukup besar untuk tabungan dan investasi, tetapi klaster yang lebih tua (Cluster 3, 4, 5) lebih memprioritaskan alokasi dana untuk tujuan jangka panjang ini.

6) Perbandingan antar metrik validasi (Davies-Bouldin Index (DBI), Silhouette Score, dan Calinski-Harabasz Index)

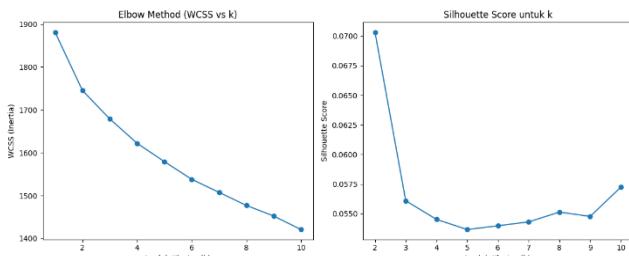


Gambar 12. Perbandingan antar metrik validasi (Davies-Bouldin Index (DBI), Silhouette Score, dan Calinski-Harabasz Index)

Berdasarkan hasil komparasi dari tiga metrik validasi (Davies-Bouldin Index, Silhouette Score, dan Calinski-Harabasz Index) diatas, berikut adalah analisis mengenai hasil yang diperoleh: Davies-Bouldin Index (DBI): DBI yang lebih kecil menunjukkan pemisahan antar klaster yang lebih baik. Nilai DBI menurun secara signifikan dari 3.56 pada k=2 ke sekitar 2.41 pada k=13 dan 15, yang menunjukkan pemisahan klaster yang semakin jelas. Untuk jumlah klaster di antara 6 hingga 10, DBI stabil di sekitar nilai 2.4, yang menunjukkan kualitas klaster yang lebih baik. Nilai DBI yang menurun seiring bertambahnya jumlah klaster menunjukkan bahwa semakin banyak klaster yang dibuat, semakin baik pemisahan antar klaster. Namun, setelah k=10, perbaikan menjadi lebih terbatas. Silhouette Score: Silhouette Score mengukur keseragaman dan jarak antar klaster. Skor yang lebih tinggi menunjukkan pemisahan yang lebih baik antara klaster-klaster yang ada. Silhouette Score sangat tinggi pada k=2, yaitu sekitar 0.070, namun menurun dengan cepat setelah k=2 dan tetap stagnan dengan nilai 0.05-0.06 di k=6 hingga k=15. Skor Silhouette menunjukkan bahwa klaster dengan k=2 memberikan pemisahan terbaik, tetapi seiring meningkatnya jumlah klaster, pemisahan antar klaster tidak meningkat signifikan. Calinski-Harabasz Index (CH Index): Calinski-Harabasz Index menilai kualitas klaster berdasarkan kekompakan dalam klaster dan pemisahan antar klaster. Semakin tinggi nilai CH, semakin baik pemisahan klaster. Nilai CH menurun drastis dari 132 pada k=2 menjadi sekitar 50-60 pada k=12 hingga k=15. Hal ini menunjukkan bahwa pemisahan antar klaster semakin menurun seiring bertambahnya jumlah klaster. Seperti DBI, nilai CH juga menunjukkan bahwa klaster antara k=6 hingga k=10 memiliki kualitas klaster yang lebih baik, meskipun penurunan skor semakin lambat pada jumlah klaster yang lebih tinggi. Kesimpulan: K=6 hingga k=10 adalah jumlah klaster yang optimal, berdasarkan analisis DBI, Silhouette Score, dan Calinski-Harabasz Index. Pada rentang ini, DBI menunjukkan pemisahan antar klaster yang jelas, Silhouette Score relatif stabil, dan nilai CH cukup tinggi. K=2 memberikan pemisahan terbaik berdasarkan Silhouette Score, namun nilai DBI dan CH menunjukkan bahwa pemisahan klaster lebih buruk dengan hanya dua klaster. Secara keseluruhan, meskipun k=6 hingga k=10 menawarkan keseimbangan terbaik dalam hal pemisahan dan pemahaman data, hasil ini masih menunjukkan bahwa lebih banyak klaster tidak memberikan peningkatan signifikan dalam kualitas klaster (terutama setelah k=10).

C. Visualisasi Hasil Klasterisasi

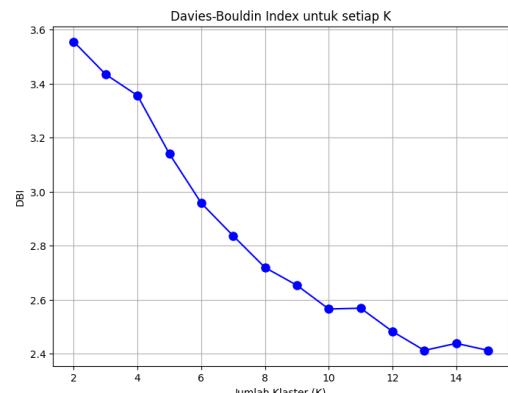
Hasil klasterisasi divisualisasikan menggunakan PCA (Principal Component Analysis) ke dalam dua dimensi untuk menggambarkan pemisahan klaster. Setiap klaster tampak menempati ruang yang berbeda, menunjukkan bahwa K-Means berhasil memisahkan kelompok Generasi Z berdasarkan gaya hidup dan prioritas finansial.



Gambar 13. Penentuan Jumlah Klaster Dengan *Elbow Method* dan *Silhouette Score*

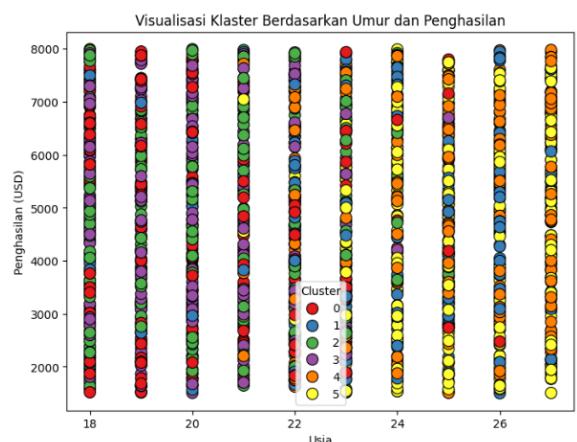
Metode Elbow Method berfokus pada penurunan WCSS (Within-Cluster Sum of Squares) untuk menentukan jumlah klaster yang optimal. Grafik menunjukkan bahwa WCSS menurun seiring bertambahnya jumlah klaster, dengan penurunan tajam terjadi antara $k=1$ hingga $k=3$, lalu melandai setelah $k=4$ hingga $k=5$. Titik siku (elbow) tampak pada $k=3$ atau $k=4$, yang menunjukkan bahwa jumlah klaster yang wajar berada di kisaran tersebut. Sementara itu, Silhouette Score mengukur seberapa baik klaster terpisah, dengan skor tertinggi ditemukan pada $k=2$ (≈ 0.07). Namun, untuk $k=3$ hingga $k=5$, skor turun drastis (≈ 0.053 – 0.056), dan pada $k=6$ hingga $k=9$, skor stagnan rendah, sedikit naik di $k=10$, tetapi tetap rendah. Hal ini menunjukkan kualitas pemisahan klaster yang lemah, meskipun dua klaster dianggap yang paling "baik" berdasarkan nilai Silhouette Score.

Berdasarkan kedua analisis tersebut, Elbow Method merekomendasikan jumlah klaster antara $k=3$ hingga $k=4$, sementara Silhouette Score menunjukkan kualitas klaster terbaik pada 2 klaster, meskipun dengan nilai rendah. Implikasi dari kedua metode ini adalah data tidak memiliki struktur klaster yang kuat, sehingga sulit memperoleh pemisahan yang jelas. Jika fokus pada kualitas pemisahan meskipun lemah, maka $k=2$ dapat dipilih. Namun, jika ingin keseimbangan antara variasi data dan efisiensi jumlah klaster, $k=3$ atau $k=4$ adalah pilihan yang lebih baik. Implikasinya, data tidak memiliki struktur klaster yang kuat, sehingga sulit mendapat pemisahan yang jelas dari Elbow Method dan Silhouette score. Namun, penggunaan DBI memberikan perspektif tambahan terhadap kualitas klaster. Berbeda dengan silhouette yang menggunakan pendekatan jarak relatif antar klaster, DBI mengukur rasio antara kedekatan dalam klaster dan jarak antar klaster, dengan nilai yang lebih kecil menunjukkan hasil klaster yang lebih baik. Dengan demikian, DBI dapat membantu menilai seberapa kompak dan terpisahnya klaster yang terbentuk, khususnya ketika nilai silhouette score rendah dan hasil Elbow Method tidak menunjukkan titik siku yang jelas.



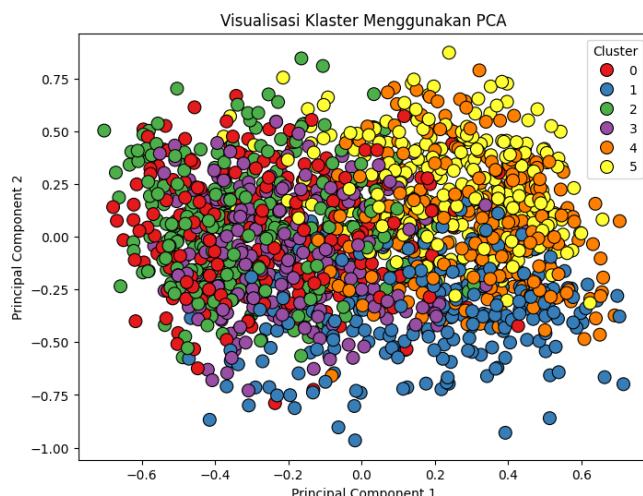
Gambar 14. Plot DBI untuk setiap K untuk visualisasi

Grafik yang ditampilkan menunjukkan hubungan antara jumlah klaster (K) dengan nilai Davies-Bouldin Index (DBI), yang digunakan untuk mengevaluasi kualitas klaster. Semakin kecil nilai DBI, semakin baik kualitas klaster yang terbentuk. Pada grafik tersebut, sumbu X (horizontal) menunjukkan jumlah klaster K yang dipilih, mulai dari 2 hingga 15, sementara sumbu Y (vertikal) menunjukkan nilai DBI untuk setiap jumlah klaster. Titik-titik pada grafik mewakili nilai DBI untuk setiap K, dan garis yang menghubungkan titik-titik ini menggambarkan tren perubahan nilai DBI seiring penambahan jumlah klaster. Analisis grafik menunjukkan bahwa nilai DBI menurun tajam saat jumlah klaster (K) bertambah dari 2 hingga sekitar 6, yang mengindikasikan bahwa penambahan jumlah klaster signifikan meningkatkan kualitas pemisahan klaster. Namun, setelah K=6, penurunan DBI mulai melambat, dan grafik cenderung stabil atau mengalami penurunan minimal di K=10 hingga K=15. Berdasarkan grafik ini, jumlah klaster optimal kemungkinan berada di antara K=6 hingga K=10, karena di rentang ini DBI menunjukkan penurunan yang stabil tanpa perubahan besar. Oleh karena itu, disarankan untuk memilih K=6 atau K=7 sebagai jumlah klaster, karena di titik tersebut DBI cukup rendah dan tidak ada penurunan besar setelahnya. Penambahan klaster lebih banyak tidak memberikan peningkatan signifikan dalam kualitas pemisahan klaster, yang terlihat dari grafik yang mulai datar setelah K=6.



Gambar 15. Visualisasi hasil klasterisasi menggunakan Scatter Plot

Grafik 15 yang ditampilkan, dapat disimpulkan bahwa data usia dan penghasilan tersebar antara 18 hingga 27 tahun, dengan penghasilan berkisar antara 1.500 hingga 8.000 USD. Tidak terlihat pola linear antara usia dengan penghasilan, yang berarti bertambahnya usia tidak selalu berbanding lurus dengan peningkatan penghasilan. Visualisasi klaster menunjukkan enam klaster (0–5) yang dibedakan berdasarkan warna, dengan setiap kelompok umur hampir selalu terwakili di semua klaster, mengindikasikan bahwa pengelompokan lebih dipengaruhi oleh kombinasi faktor usia dan penghasilan, bukan hanya salah satunya. Beberapa klaster, seperti Klaster 0 (merah) dan Klaster 2 (hijau), memiliki distribusi yang relatif merata di semua kelompok umur, sementara Klaster 4 (orange) dan Klaster 5 (kuning) lebih dominan pada usia yang lebih tua (24–27) dengan rentang penghasilan yang bervariasi. Kesimpulannya, meskipun usia yang sama bisa masuk ke klaster yang berbeda, analisis clustering ini sangat berguna untuk memahami bahwa meskipun usia serupa, ada perbedaan karakteristik penghasilan yang mungkin dipengaruhi oleh faktor lain seperti pekerjaan, pendidikan, atau pengalaman. Hal ini sejalan dan didukung dengan visualisasi Scatter Plot PCA sebagai berikut.



Gambar 16. Visualisasi hasil klasterisasi menggunakan Scatter Plot PCA

Pada scatter plot PCA yang ditampilkan, kita dapat mengamati pemisahan klaster berdasarkan dua komponen utama (Principal Component 1 dan Principal Component 2). Setiap titik mewakili individu dalam dataset, dan warna masing-masing titik mengindikasikan klaster yang ditentukan oleh algoritma K-Means. Secara umum, scatter plot menunjukkan pemisahan yang cukup jelas antar klaster, meskipun ada beberapa area di mana titik dari klaster yang berbeda mungkin sedikit tumpang tindih. Ini menunjukkan bahwa meskipun ada pemisahan yang baik antara klaster, masih ada beberapa kemiripan dalam karakteristik pengeluaran di beberapa klaster. Klaster 0 (Merah): Terlihat agak tersebar di sisi kiri bawah grafik. Titik-titik dari klaster ini menunjukkan variasi yang cukup besar di sekitar dua komponen utama, tetapi masih bisa dibedakan dari klaster lain. Klaster 1 (Biru): Sebagian besar titik pada klaster ini

tersebar di bagian tengah, dengan konsentrasi lebih besar pada sumbu horizontal (Principal Component 1) yang sedikit lebih tinggi. Klaster 2 (Hijau): Klaster ini lebih tersebar di sisi kanan atas, menunjukkan bahwa individu dalam klaster ini memiliki karakteristik pengeluaran yang lebih tinggi atau berbeda dalam komponen utama. Klaster 3 (Kuning): Dapat dilihat bahwa klaster ini lebih terpusat dan sedikit terpisah dari klaster lain. Klaster 4 (Orange) dan Klaster 5 (Ungu): Kedua klaster ini menunjukkan distribusi yang lebih menyebar, tetapi masih ada pemisahan yang dapat diidentifikasi dengan jelas. Klaster 5, khususnya, memiliki pemisahan yang sedikit lebih rapat. Interpretasi Sumbu PCA:

Principal Component 1 (PCA1) merupakan komponen utama yang menjelaskan sebagian besar variasi dalam data. Klaster yang lebih terpisah pada sumbu ini menunjukkan perbedaan signifikan dalam pengeluaran untuk kategori tertentu, seperti hiburan, belanja online, atau tabungan. Principal Component 2 (PCA2) mungkin lebih berfokus pada variasi dalam pengeluaran untuk kebutuhan dasar atau pengelolaan keuangan jangka panjang, yang mempengaruhi klaster yang lebih dewasa dan lebih terencana.

Visualisasi ini memperlihatkan bahwa klaster yang terbentuk oleh K-Means relatif sesuai dengan kategori yang diinginkan, meskipun ada sedikit tumpang tindih di antara beberapa klaster. Ini mungkin menunjukkan bahwa perilaku konsumsi individu memiliki kesamaan di beberapa aspek, meskipun klaster-klaster ini memiliki prioritas yang berbeda dalam pengeluaran. Kesimpulan secara keseluruhan, scatter plot PCA menunjukkan bahwa analisis K-Means berhasil mengidentifikasi pola perilaku pengeluaran yang berbeda di antara generasi Z, dengan pemisahan yang cukup jelas meskipun terdapat sedikit tumpang tindih. Beberapa klaster, terutama klaster dengan usia lebih muda (seperti Klaster 0 dan 2), menunjukkan kecenderungan konsumtif yang lebih tinggi (hiburan, belanja online), sementara klaster yang lebih tua (seperti Klaster 3, 4, dan 5) lebih fokus pada tabungan dan investasi. Analisis ini juga memiliki implikasi praktis, seperti segmentasi pasar, program pengembangan karier, dan penetapan kebijakan ekonomi yang lebih tepat sasaran berdasarkan usia dan penghasilan. Interpretasi per klaster adalah sebagai berikut :

Klaster 0 (Merah) terdiri dari individu Generasi Z dengan usia antara 18 hingga 27 tahun, yang memiliki penghasilan yang bervariasi antara 1.500 hingga 8.000 USD, menciptakan klaster yang bersifat heterogen. Pola pengeluaran mereka mencakup kebutuhan dasar seperti tempat tinggal dan hiburan, dengan variasi besar dalam pengeluaran untuk tabungan dan investasi. Dari sisi kebijakan, klaster ini memerlukan pendekatan yang inklusif dan umum dalam pendidikan keuangan, mengingat ragam penghasilan dan kondisi sosial-ekonomi yang ada. Bisnis dapat merancang produk finansial yang fleksibel, seperti aplikasi keuangan yang dapat membantu pengelolaan tabungan dan pengeluaran dasar untuk kelompok ini.

Klaster 1 (Biru) terdiri dari individu yang dominan berusia antara 24 hingga 27 tahun, dengan penghasilan di atas 4.000 USD hingga 8.000 USD, menunjukkan kelompok yang lebih

mapan, mungkin karena pengalaman kerja yang lebih lama atau pekerjaan tetap. Pengeluaran mereka lebih terfokus pada kebutuhan dasar dan tabungan, dengan prioritas pada investasi. Kebijakan yang tepat untuk klaster ini adalah program perencanaan keuangan dan pensiun untuk membantu mereka merencanakan masa depan keuangan mereka. Bisnis di sektor asuransi, investasi, dan perencanaan keuangan dapat mengembangkan produk yang berfokus pada pengelolaan keuangan jangka panjang untuk memenuhi kebutuhan klaster ini.

Klaster 2 (Hijau) menunjukkan distribusi usia yang merata antara 18 hingga 27 tahun, dengan penghasilan mayoritas berada di rentang 3.000 hingga 6.000 USD, menggambarkan kelompok dengan penghasilan menengah. Pola pengeluaran mereka seimbang antara kebutuhan dasar, hiburan, dan tabungan serta investasi. Untuk kebijakan, program peningkatan literasi keuangan yang mendorong pengelolaan tabungan dan investasi dapat membantu mereka lebih terlibat dalam perencanaan finansial jangka panjang. Bisnis dapat menawarkan produk yang memberikan insentif tabungan atau investasi jangka panjang, karena mereka memiliki keseimbangan antara konsumsi dan perencanaan keuangan.

Klaster 3 (Ungu) didominasi oleh individu berusia 18 hingga 22 tahun dengan sebagian besar penghasilan mereka di bawah 5.000 USD, menggambarkan tahap awal karier atau status mahasiswa. Pengeluaran mereka lebih cenderung untuk kebutuhan dasar dan hiburan, dengan sedikit pengeluaran untuk tabungan dan investasi. Untuk klaster ini, kebijakan yang mendukung pendidikan keuangan dasar dan literasi keuangan yang lebih kuat sangat diperlukan untuk membantu mereka mengelola pengeluaran dan mulai menabung untuk masa depan. Bisnis dapat mengembangkan produk dengan pendekatan micro-investing atau tabungan berbasis gamifikasi, memungkinkan mereka mulai berinvestasi meskipun penghasilan mereka terbatas.

Klaster 4 (Oranye) lebih banyak terdiri dari individu usia 24 tahun ke atas, dengan penghasilan lebih dari 5.000 USD, lebih dominan pada rentang atas. Mereka lebih fokus pada kebutuhan dasar dan investasi, dengan pengeluaran moderat untuk hiburan. Kebijakan yang mengedepankan perencanaan keuangan jangka panjang dan investasi untuk pensiun akan sangat relevan bagi mereka. Selain itu, kebijakan yang mendorong mereka untuk berinvestasi dalam produk keuangan berisiko rendah juga dapat meningkatkan ketstabilan finansial mereka. Penyedia produk pensiun, investasi, dan asuransi dapat menawarkan produk yang lebih matang dan berorientasi pada perencanaan masa depan, seperti investasi real estate atau portofolio saham jangka panjang.

Klaster 5 (Kuning) terdiri dari individu dengan usia antara 24 hingga 27 tahun dan penghasilan yang stabil di rentang 1.500 hingga 4.000 USD. Mereka cenderung mengalokasikan lebih banyak dana untuk tabungan dan investasi, dengan pengeluaran moderat untuk kebutuhan dasar dan hiburan. Kebijakan yang fokus pada pengelolaan tabungan dan penghasilan menengah sangat bermanfaat bagi mereka, dengan memperkenalkan produk yang meningkatkan

stabilitas keuangan mereka. Penyedia produk investasi yang fleksibel, seperti investasi mikro atau funds kecil, akan diterima dengan baik oleh mereka yang ingin mengoptimalkan penghasilan terbatas mereka.

Kesimpulannya, individu berusia 18–22 tahun lebih sering masuk ke dalam Klaster 3, yang memiliki penghasilan rendah hingga menengah. Pada usia transisi (23–24 tahun), terdapat variasi, dengan beberapa mulai masuk ke Klaster 4 dan Klaster 5. Sedangkan pada usia yang lebih matang (25–27 tahun), lebih banyak individu yang masuk ke Klaster 1, Klaster 4, dan Klaster 5, yang memiliki penghasilan menengah ke atas. Jadi, klasterisasi ini menunjukkan adanya pola perkembangan karier: semakin bertambah usia, peluang untuk masuk klaster dengan penghasilan menengah-atas semakin besar, meskipun ada sebagian yang masih bertahan di kelompok penghasilan rendah.

D. Implikasi Hasil Penelitian

Hasil penelitian ini memberikan wawasan yang berharga bagi pelaku bisnis dan lembaga keuangan untuk merancang strategi pemasaran yang lebih terpersonalisasi bagi Generasi Z. Lembaga keuangan dan fintech juga dapat memanfaatkan temuan ini untuk mengembangkan produk keuangan yang lebih terpersonalisasi, sesuai dengan klaster perilaku yang teridentifikasi. Hasil penelitian ini juga dapat digunakan oleh pemerintah dan pendidikan untuk merancang strategi literasi keuangan berdasarkan usia dan gaya hidup. Bagi perusahaan, ini membuka peluang untuk menyusun strategi pemasaran produk yang lebih tepat sasaran untuk Generasi Z.

Terakhir, penelitian ini memberikan kontribusi penting dalam bidang machine learning dan analisis perilaku konsumen, khususnya dengan penggunaan algoritma K-Means untuk segmentasi perilaku finansial Generation Z. Melalui penerapan metrik evaluasi klaster yang lebih komprehensif, seperti Davies-Bouldin Index (DBI), Silhouette Score, dan Calinski-Harabasz Index, penelitian ini memperkaya pendekatan tradisional dalam klasterisasi data finansial, yang selama ini lebih fokus pada pemisahan yang jelas antar klaster tanpa mempertimbangkan kualitas dan keseimbangan dalam klasterisasi. Dalam konteks consumer analytics, studi ini membuka peluang untuk pemahaman lebih dalam mengenai pola pengeluaran konsumen muda, yang dapat dijadikan dasar untuk penelitian lebih lanjut dalam merancang model-model prediktif yang lebih akurat. Oleh karena itu, temuan ini tidak hanya relevan untuk aplikasi bisnis, tetapi juga memberikan kontribusi yang signifikan dalam mengembangkan teori perilaku konsumen yang berbasis data dan penerapannya dalam pengambilan keputusan berbasis data di bidang keuangan digital dan pemasaran.

IV. KESIMPULAN

Penelitian ini memberikan kontribusi baru dengan menganalisis dataset "Gen Z Money Spending" dari Kaggle secara mendalam, yang belum banyak dieksplorasi sebelumnya dalam konteks perilaku keuangan. Penelitian ini juga memperkenalkan penggunaan algoritma K-Means untuk segmentasi yang lebih baik berdasarkan pola pengeluaran Generasi Z, serta memperkenalkan penggunaan metrik evaluasi klaster yang lebih komprehensif, seperti Davies-Bouldin Index, Silhouette Coefficient, dan Calinski-Harabasz Index, untuk memastikan kualitas klaster yang lebih akurat.

Dengan pendekatan ini, penelitian ini tidak hanya memberikan pemahaman lebih dalam tentang perilaku konsumsi dan keuangan Generasi Z, tetapi juga memberikan wawasan yang lebih terperinci mengenai segmentasi yang lebih efektif dan relevansi penggunaan algoritma dan metrik evaluasi dalam analisis klasterisasi. Berdasarkan hasil penelitian, Penelitian ini menunjukkan bahwa Generasi Z memiliki pola pengeluaran yang sangat bervariasi, yang dipengaruhi oleh usia, penghasilan, gaya hidup, dan kebutuhan pribadi. Dengan menggunakan algoritma K-Means, perilaku belanja mereka terbagi dalam enam klaster yang berbeda, mulai dari kelompok dengan gaya hidup konsumtif yang fokus pada hiburan dan belanja online, hingga kelompok dengan prioritas stabil pada pengeluaran untuk kebutuhan dasar, tabungan, dan investasi. Evaluasi menggunakan Davies-Bouldin Index (DBI) menunjukkan bahwa jumlah klaster optimal terletak di antara $k=6$ hingga $k=10$, dengan pemisahan antar klaster yang baik. Hasil segmentasi ini memberikan wawasan penting bagi bisnis, penyedia layanan keuangan, dan pengembang produk untuk merancang strategi pemasaran yang lebih personal dan sesuai dengan karakteristik masing-masing kelompok.

DAFTAR PUSTAKA

- [1] L. Sekar Arum, Amira Zahrani, and N. A. Duha, "Karakteristik Generasi Z dan Kesiapannya dalam Menghadapi Bonus Demografi 2030," *Account. Student Res. J.*, vol. 2, no. 1, pp. 59–72, Mar. 2023, doi: 10.62108/asrj.v2i1.5812.
- [2] Muhammad Adnan Faidh, Muhamad Esa Maulana, Ninda Ela Putri, Siti Indriyani Putri, Thasya Azhari Munir, and April Laksana, "Peran Media Sosial X Dalam Perkembangan Komunikasi Di Era Digital," *Konsensus J. Ilmu Pertahanan, Huk. dan Ilmu Komun.*, vol. 1, no. 6, pp. 43–51, 2024, doi: 10.62383/konsensus.v1i6.433.
- [3] A. Jordan and K. Nuringsih, "Understanding Financial Behavior in Generation Z," *Int. J. Appl. Econ. Bus.*, vol. 1, no. 4, pp. 2535–2546, 2023, doi: 10.24912/ijaeb.v1i4.2535-2546.
- [4] S. Mu and J. Jurana, "Financial Behavior Patterns Of Generation Z : Netnographic Analysis Of The Fear Of Missing Out (Fomo) Phenomenon," pp. 23–34, 2025.
- [5] K. Tabanian, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustainability*, vol. 14, no. 12, p. 7243, Jun. 2022, doi: 10.3390/su14127243.
- [6] F. Akibun, H. Prayitno, R. A. Z., and N. M. Otto, "Financial Literacy In Gen Z Generation (Case Study at Bina Taruna University Gorontalo)," no. 2, pp. 1–8, 2025.
- [7] J. M. Rodriguez, "The Mediation of Financial Behavior to Financial Literacy and Spending Habits of Gen Z : An Exploratory Factor Analysis," vol. 5, no. 2, 2024.
- [8] J. Chitra and J. Heikal, "Customer segmentation using the K-Means Clustering algorithm in Foreign Banks in Indonesia," *Indones. Account. Res. J.*, vol. 11, no. 4, pp. 230–241, 2024.
- [9] N. Jain and V. Ahuja, "Segmenting online consumers using K-means cluster analysis," *Int. J. Logist. Econ. Glob.*, vol. 6, no. 2, p. 161, 2014, doi: 10.1504/IJLEG.2014.068274.
- [10] Z. Liu, Y. Li, C. Liu, X. Zhao, and W. Yin, "Application of K-Means Clustering Algorithm in Analyzing College Students' Mental Health," in *2024 3rd International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS)*, IEEE, Jul. 2024, pp. 175–180, doi: 10.1109/AIARS63200.2024.00038.
- [11] E. Omol, D. Onyangor, L. Mburu, and P. Abuonji, "Application Of K-Means Clustering For Customer Segmentation In Grocery Stores In Kenya," *Int. J. Sci. Technol. Manag.*, vol. 5, no. 1, pp. 192–200, Jan. 2024, doi: 10.46729/ijstm.v5i1.1024.
- [12] Z. Zhu and N. Liu, "Early Warning of Financial Risk Based on K-Means Clustering Algorithm," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/5571683.
- [13] C. Wongoutong, "The impact of neglecting feature scaling in k-means clustering," *PLoS One*, vol. 19, no. 12, p. e0310839, Dec. 2024, doi: 10.1371/journal.pone.0310839.
- [14] M. Prasad and S. T, "Clustering Accuracy Improvement Using Modified Min-Max Normalization Technique," Nov. 07, 2024, doi: 10.20944/preprints202411.0486.v1.
- [15] K. D. Tran, H. T. Phan, C. T. K. Nguyen, B. C. Nguyen, H. V. Le, and K. D. Nguyen, "Solvent-Free Synthesis of Co-Based Zeolitic Imidazolate Framework (ZIF-9) for the Removal of Congo Red from Water," *Indones. J. Chem.*, vol. 25, no. 1, p. 178, Jan. 2025, doi: 10.22146/ijc.99141.
- [16] E. Sulaiman, N. Nopriyeni, C. Darwin, and A. Lusianti, "Diversity Of Liana Plants Available In The Konak Protected Forest Area, Kepahiang District, Kepahiang Regency," *J. Pembelajaran dan Biol. Nukl.*, vol. 8, no. 3, pp. 820–830, Nov. 2022, doi: 10.36987/jpbn.v8i3.3170.
- [17] A. R. F. Falih, R. Kurniawan, Y. Arie Wijaya, and S. Anwar, "Algoritma K-Mean Untuk Optimalisasi Model Clustering Data Penjualan Toko Online Di Tiktok Shop Dalam Strategi Pemasaran," *J. Sist. Inf. Kaputama*, vol. 9, no. 1, pp. 1–11, 2025, doi: 10.59697/jsik.v9i1.929.
- [18] M. Zhulal, S. A. Marits, and S. Herman, "Generation Z Purchasing Behavior Profile in the Digital Economy: Normative Analysis in Online Markets," *J. Ilm. Manaj. Kesatuan*, vol. 12, no. 1, pp. 1–8, 2024, doi: 10.37641/jimkes.v12i1.2326.
- [19] N. Sureja, B. Chawda, and A. Vasant, "An improved K-medoids clustering approach based on the crow search algorithm," *J. Comput. Math. Data Sci.*, vol. 3, p. 100034, Jun. 2022, doi: 10.1016/j.jcmds.2022.100034.
- [20] A. Gere, "Recommendations for validating hierarchical clustering in consumer sensory projects," *Curr. Res. Food Sci.*, vol. 6, p. 100522, 2023, doi: 10.1016/j.crefs.2023.100522.
- [21] M. Hahsler, M. Piekenbrock, and D. Doran, "dbscan : Fast Density-Based Clustering with R," *J. Stat. Softw.*, vol. 91, no. 1, 2019, doi: 10.18637/jss.v091.i01.
- [22] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (Ny.)*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/j.ins.2022.11.139.
- [23] M. Zubair, M. A. Iqbal, A. Shil, M. J. M. Chowdhury, M. A. Moni, and I. H. Sarker, "An Improved K-means Clustering Algorithm Towards an Efficient Data-Driven Modeling," *Ann. Data Sci.*, vol. 11, no. 5, pp. 1525–1544, Oct. 2024, doi: 10.1007/s40745-022-00428-2.
- [24] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," *Electronics*, vol. 9, no. 8, p. 1295, Aug. 2020, doi: 10.3390/electronics9081295.
- [25] K. N. Salam, A. W. T. F. Singkeruang, M. F. Husni, B. Baharuddin, and D. P. A.R, "Gen-Z Marketing Strategies: Understanding Consumer Preferences and Building Sustainable Relationships,"

- [26] *Golden Ratio Mapp. Idea Lit. Format*, vol. 4, no. 1, pp. 53–77, 2024, doi: 10.52970/grmilf.v4i1.351.
- [27] E. U. Oti, M. O. Olusola, F. C. Eze, and S. U. Enogwe, “Comprehensive Review of K-Means Clustering Algorithms,” *Int. J. Adv. Sci. Res. Eng.*, vol. 07, no. 08, pp. 64–69, 2021, doi: 10.31695/IJASRE.2021.34050.
- [28] D. A. Tarigan, “Optimization of the K-Means Clustering Algorithm Using Davies Bouldin Index in Iris Data Classification,” *Media Online*, vol. 4, no. 1, pp. 545–552, 2023, doi: 10.30865/klik.v4i1.964.
- [29] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.
- H. Hassani, M. Kalantari, and C. Beneki, “Comparative Assessment of Hierarchical Clustering Methods for Grouping in Singular Spectrum Analysis,” *AppliedMath*, vol. 1, no. 1, pp. 18–36, Dec. 2021, doi: 10.3390/appliedmath1010003.