

Segmentasi Konsumen *E-Commerce* Menggunakan Algoritma K-Means Berbasis *Dashboard* Interaktif

Aru Chevy Rasyid¹, Fauziah^{2*}

¹ *Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional, Pasar Minggu, Jakarta Selatan, Indonesia*
E-mail: aruchvyrasyid@gmail.com

² *Universitas Nasional, Pasar Minggu, Jakarta Selatan, Indonesia*
E-mail: fauziah@civitas.unas.ac.id

*Corresponding Author

Abstrak— Perkembangan industri *e-commerce* di Indonesia mendorong peningkatan volume data perilaku konsumen yang signifikan. Kondisi ini menuntut adanya pendekatan analitik yang mampu menggambarkan pola transaksi secara lebih terstruktur guna mendukung proses segmentasi konsumen. Penelitian ini bertujuan untuk mengelompokkan konsumen *e-commerce* menggunakan metode K-Means Clustering serta memvisualisasikan hasil analisis dalam bentuk dashboard interaktif. Proses penelitian diawali dengan pembersihan dan standarisasi data, kemudian dilanjutkan dengan pembentukan variabel turunan berupa average basket, spend per month, dan paylater share sebagai representasi perilaku transaksi konsumen. Penentuan jumlah kluster optimal dilakukan melalui evaluasi beberapa metrik, yaitu grafik Sum of Squared Errors (SSE), Silhouette Score, Calinski–Harabasz Index, dan Davies–Bouldin Index. Meskipun nilai Silhouette Score dan Calinski–Harabasz Index tertinggi diperoleh pada $k = 2$, hasil evaluasi secara keseluruhan menunjukkan bahwa $k = 3$ memberikan struktur kluster yang lebih seimbang dan mudah diinterpretasikan. Berdasarkan hasil tersebut, terbentuk tiga segmen konsumen, yaitu konsumen dengan aktivitas belanja menengah, konsumen bernilai tinggi dengan kecenderungan penggunaan Paylater, serta konsumen dengan frekuensi transaksi tinggi namun nilai pembelian relatif rendah. Seluruh hasil analisis kemudian diimplementasikan dalam dashboard interaktif berbasis Streamlit untuk mendukung pengambilan keputusan berbasis data.

Kata Kunci — Analisis Data, Dashboard Interaktif, E-commerce, K-Means Clustering, Segmentasi Konsumen.

Abstract— The rapid expansion of the *e-commerce* industry in Indonesia has led to a substantial increase in consumer behavioral data. This condition highlights the need for analytical approaches capable of identifying transaction patterns and supporting meaningful customer segmentation. This study aims to segment *e-commerce* customers using the K-Means clustering method and to present the results through an interactive dashboard. The research process includes data cleaning and standardization, followed by the construction of derived variables such as average basket, spend per month, and paylater share to represent customer transaction behavior. The optimal number of clusters was determined by evaluating several metrics, including the Sum of Squared Errors (SSE) graph, Silhouette Score, Calinski–Harabasz Index, and Davies–Bouldin Index. Although the highest Silhouette Score and Calinski–Harabasz Index were obtained at $k = 2$, the overall evaluation indicated that $k = 3$ produced a more balanced and interpretable clustering structure. Consequently, three customer segments were identified: customers with moderate shopping activity, high-value customers with a tendency to use Paylater services, and customers with high transaction frequency but relatively low purchase value. The analytical results were subsequently implemented in a Streamlit-based interactive dashboard to support data-driven decision making.

Keywords — Customer Segmentation, Data Analysis, E-commerce, Interactive Dashboard, K-Means Clustering.

I. PENDAHULUAN

Perkembangan ekonomi digital di Indonesia mendorong pertumbuhan sektor *e-commerce* secara signifikan dalam beberapa tahun terakhir, seiring meningkatnya penetrasi internet, adopsi perangkat digital, serta perubahan pola konsumsi masyarakat. Peningkatan aktivitas transaksi *e-commerce* tercermin dari bertambahnya volume dan nilai transaksi yang dihasilkan oleh konsumen lintas wilayah, termasuk wilayah tier 2 dan tier 3, yang menunjukkan semakin meratanya akses terhadap layanan digital. Perluasan basis pengguna ini menghasilkan data transaksi dengan karakteristik perilaku yang semakin beragam, baik dari sisi frekuensi transaksi, nilai pembelian, maupun preferensi metode pembayaran digital, termasuk layanan *buy now pay later* (BNPL) atau *Paylater* [1], [2].

Meningkatnya volume dan kompleksitas data transaksi tersebut menuntut pemanfaatan pendekatan analisis berbasis data untuk memahami perilaku konsumen secara lebih mendalam. Segmentasi konsumen berbasis perilaku transaksi menjadi pendekatan yang relevan karena mampu mengelompokkan konsumen ke dalam segmen yang lebih homogen berdasarkan pola pembelian aktual. Sejumlah penelitian menunjukkan bahwa segmentasi berbasis data transaksi lebih representatif dibandingkan segmentasi demografis karena mencerminkan perilaku aktual konsumen dalam berinteraksi dengan platform *e-commerce* [3], [4].

Berbagai penelitian terdahulu telah menerapkan algoritma K-Means *Clustering* untuk melakukan segmentasi konsumen *e-commerce* berdasarkan perilaku pembelian. Contoh dari artikel pada jurnal dalam [3] menunjukkan bahwa K-Means efektif dalam mengelompokkan konsumen berdasarkan kesamaan pola pengeluaran dan intensitas transaksi, sementara artikel pada jurnal dalam [5] membuktikan bahwa metode ini mampu menghasilkan segmentasi yang mendukung pengambilan keputusan pemasaran berbasis data transaksi penjualan. Penelitian lain juga menegaskan bahwa K-Means memiliki keunggulan dari sisi kesederhanaan, efisiensi komputasi, dan interpretabilitas hasil klaster [6], [7], [8].

Untuk meningkatkan kualitas segmentasi, sejumlah penelitian menggabungkan K-Means dengan pendekatan *feature engineering*, khususnya model RFM. Artikel pada jurnal dalam [9] menunjukkan bahwa integrasi RFM dengan K-Means menghasilkan segmentasi yang lebih stabil dan interpretatif pada data transaksi *e-commerce*. Pendekatan serupa juga diterapkan oleh artikel pada jurnal dalam [10], [11], yang menegaskan bahwa pemilihan fitur berbasis perilaku transaksi berperan penting dalam meningkatkan kualitas klaster yang dihasilkan.

Di sisi lain, kajian mengenai penggunaan layanan *Paylater* dalam konteks *e-commerce* menunjukkan bahwa fasilitas pembayaran ini memiliki pengaruh terhadap perilaku pembelian konsumen. Dataset dan analisis yang dikembangkan oleh artikel pada jurnal dalam [1] mengindikasikan bahwa penggunaan *Paylater* berkorelasi dengan kecenderungan pembelian impulsif dan peningkatan intensitas transaksi. Namun, sebagian besar penelitian tersebut masih berfokus pada analisis perilaku berbasis survei atau model persepsi, sehingga pemanfaatan data transaksi aktual dalam segmentasi konsumen berbasis *Paylater* masih relatif terbatas.

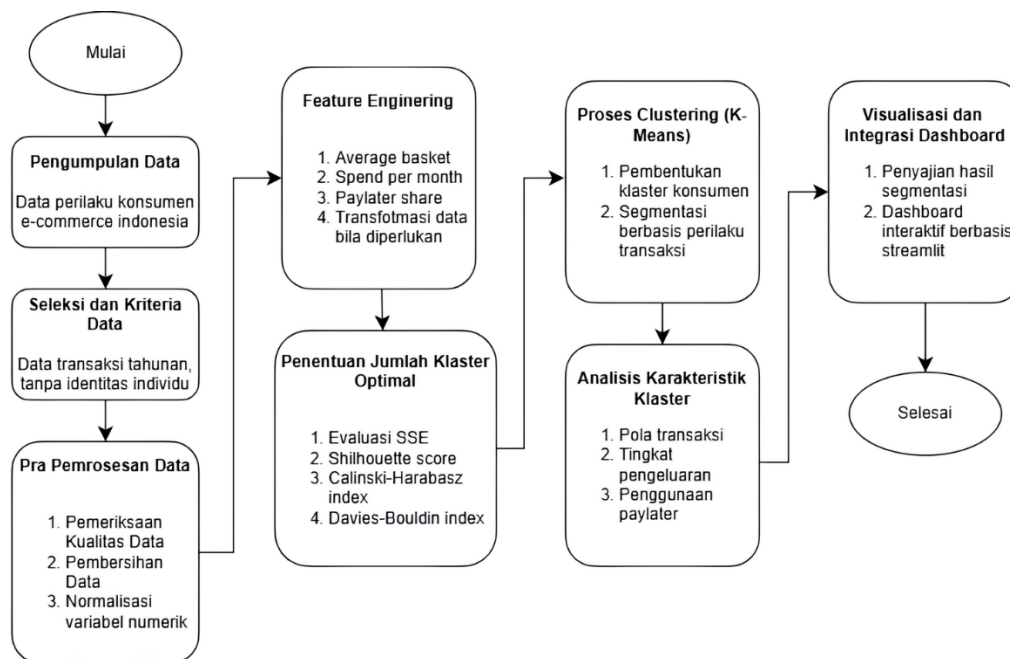
Selain aspek metodologis, keterbatasan juga ditemukan pada penyajian hasil segmentasi. Sebagian besar penelitian menyajikan hasil *clustering* dalam bentuk visualisasi statis, yang kurang mendukung eksplorasi data secara dinamis oleh pengambil keputusan. Dalam jurnal [12] menekankan bahwa integrasi analitik ke dalam *dashboard* interaktif dapat meningkatkan pemahaman pengguna terhadap data dan mendukung pengambilan keputusan berbasis bukti secara lebih efektif [13].

Berdasarkan tinjauan terhadap penelitian terdahulu, dapat diidentifikasi celah penelitian pada tiga aspek utama, yaitu pemanfaatan data transaksi aktual untuk segmentasi perilaku konsumen, integrasi indikator *Paylater* sebagai bagian dari *feature engineering*, serta penyajian hasil segmentasi dalam bentuk *dashboard* interaktif yang mudah digunakan. Hingga saat ini, penelitian yang menggabungkan ketiga aspek tersebut dalam satu kerangka analisis terpadu masih relatif terbatas.

Oleh karena itu, penelitian ini bertujuan untuk menerapkan metode K-Means *Clustering* dalam melakukan segmentasi perilaku konsumen *e-commerce* di Indonesia dengan memanfaatkan variabel transaksi hasil *feature engineering*, termasuk indikator penggunaan *Paylater*, serta mengintegrasikan hasil segmentasi ke dalam *dashboard* interaktif berbasis *Streamlit*. *State of the art* penelitian ini terletak pada penggabungan pendekatan segmentasi berbasis data transaksi aktual, analisis perilaku *Paylater*, dan visualisasi interaktif dalam satu sistem pendukung keputusan. Berdasarkan tujuan tersebut, hipotesis penelitian ini adalah bahwa penerapan K-Means *Clustering* dengan *feature engineering* yang tepat mampu menghasilkan segmentasi konsumen yang bermakna secara perilaku dan dapat disajikan secara efektif melalui *dashboard* interaktif.

II. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan teknik *data mining* untuk menganalisis dan mengelompokkan perilaku konsumen *e-commerce* di Indonesia. Pendekatan ini dipilih karena mampu mengidentifikasi pola perilaku transaksi berdasarkan data numerik dalam skala besar serta menghasilkan segmentasi konsumen yang bersifat objektif dan berbasis data aktual [3], [14].



Gambar 1. Diagram Alur Penelitian

Objek penelitian berupa data perilaku konsumen *e-commerce* di Indonesia yang bersumber dari *Laporan Perilaku Konsumen E-commerce Indonesia 2023* yang diterbitkan oleh Kredivo dan Katadata Insight Center. Data yang digunakan merupakan data sekunder bersifat sintesis, dikembangkan dari agregasi transaksi pengguna *e-commerce* tanpa memuat identitas individu, sehingga tetap menjaga aspek keamanan dan kerahasiaan data [15]. Kriteria inklusi data mencakup konsumen yang memiliki aktivitas transaksi selama satu periode tahunan dan tercatat pada variabel transaksi serta metode pembayaran. Variabel terikat dalam penelitian ini adalah hasil segmentasi konsumen, sedangkan variabel bebas terdiri atas variabel perilaku transaksi yang merepresentasikan intensitas dan nilai belanja konsumen. Variabel utama yang digunakan meliputi frekuensi transaksi, total pengeluaran, serta proporsi penggunaan layanan *Paylater*. Untuk memperkuat representasi perilaku konsumen, dilakukan pembentukan variabel turunan (*feature engineering*), yaitu *average basket*, *spend per month*, dan *paylater share*. Pendekatan ini umum digunakan dalam penelitian segmentasi konsumen berbasis perilaku karena mampu meningkatkan kejelasan perbedaan antar segmen [10], [11], [16].

Sebelum proses segmentasi, data melalui tahapan pra-pemrosesan yang meliputi pemeriksaan kualitas data, pembersihan, dan standarisasi. Seluruh variabel numerik dinormalisasi untuk menghindari dominasi variabel dengan rentang nilai besar dalam perhitungan jarak, sebagaimana direkomendasikan dalam penelitian *clustering* berbasis transaksi *e-commerce* [14]. Selain itu, dilakukan transformasi logaritmik pada variabel transaksi tertentu untuk mengatasi distribusi data yang bersifat tidak simetris dan meningkatkan stabilitas hasil analisis [17].

Penentuan jumlah kluster dilakukan melalui evaluasi multi-metrik, yaitu *Sum of Squared Errors (SSE)*, *Silhouette Score*, *Calinski–Harabasz Index*, dan *Davies–Bouldin Index*. Penggunaan beberapa metrik evaluasi secara simultan bertujuan memperoleh struktur kluster yang lebih seimbang dan objektif, sehingga pemilihan jumlah kluster tidak bergantung pada satu indikator saja [10], [18], [19]. Berdasarkan hasil evaluasi tersebut, proses segmentasi konsumen dilakukan menggunakan metode *K-Means Clustering* untuk membentuk kelompok konsumen dengan karakteristik perilaku yang relatif homogen. Hasil klusterisasi kemudian dianalisis untuk mengidentifikasi karakteristik utama setiap segmen, mencakup pola transaksi, tingkat pengeluaran, serta kecenderungan penggunaan layanan *Paylater*.

Tahap akhir penelitian adalah analisis dan visualisasi hasil, di mana seluruh *output* segmentasi diintegrasikan ke dalam *dashboard* interaktif berbasis *streamlit*. *Dashboard* ini dirancang sebagai alat eksplorasi data dan pendukung pengambilan keputusan berbasis data, dengan menampilkan visualisasi kluster, statistik perilaku, serta analisis lanjutan secara interaktif. Pendekatan integrasi analitik dan

visualisasi ini sejalan dengan penelitian sebelumnya yang menekankan peran *dashboard* interaktif dalam meningkatkan interpretasi hasil segmentasi konsumen [12], [20].

III. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian dan pembahasan yang mencakup perancangan aplikasi, implementasi sistem, serta analisis hasil segmentasi perilaku konsumen *e-commerce* di Indonesia. Bagian awal bab difokuskan pada penyajian rancangan dan tampilan aplikasi *dashboard* interaktif sebagai media penyajian hasil analisis, diikuti dengan deskripsi dataset, analisis eksploratif data, penentuan jumlah kluster optimal, serta hasil penerapan metode K-Means *Clustering*. Pembahasan pada setiap tahapan disajikan secara terstruktur dengan merujuk pada tabel dan gambar yang relevan untuk memperjelas hasil yang diperoleh dan mendukung interpretasi temuan penelitian.

Rancangan Aplikasi dan *Layout* Sistem

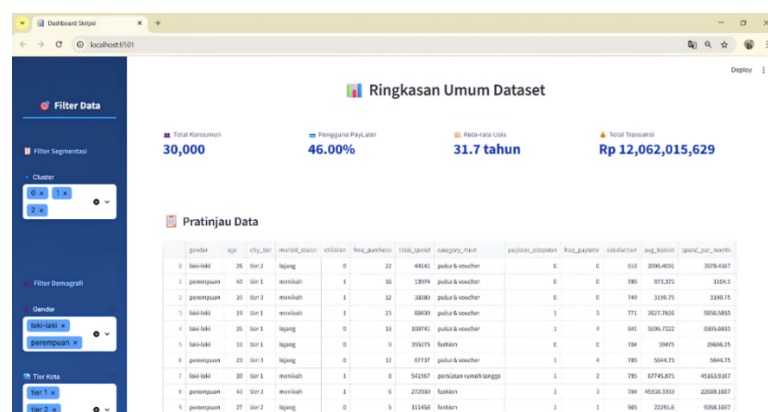
Penelitian ini menghasilkan sebuah aplikasi *dashboard* interaktif berbasis *streamlit* yang dirancang untuk menampilkan hasil segmentasi perilaku konsumen *e-commerce* secara visual dan mudah dipahami. Aplikasi ini menyediakan antarmuka *input* dan *output* yang memungkinkan pengguna untuk mengeksplorasi karakteristik konsumen berdasarkan hasil *clustering*, distribusi variabel transaksi, serta penggunaan layanan *Paylater*.

Rancangan *layout* aplikasi terdiri dari beberapa komponen utama, yaitu ringkasan informasi kluster, visualisasi distribusi konsumen, serta tampilan analisis perilaku transaksi. Tampilan awal aplikasi dirancang untuk memberikan gambaran umum hasil segmentasi sebelum pengguna melakukan eksplorasi lebih lanjut.

Gambar 2 menampilkan ringkasan umum dataset konsumen *e-commerce* yang digunakan dalam penelitian ini. Informasi agregat yang disajikan mencakup jumlah total konsumen, proporsi pengguna layanan *PayLater*, rata-rata usia konsumen, serta total nilai transaksi selama periode pengamatan. Penyajian ringkasan ini bertujuan untuk memberikan gambaran awal mengenai skala data dan karakteristik umum konsumen sebelum dilakukan analisis lebih lanjut.

Proporsi pengguna *PayLater* yang relatif besar menunjukkan bahwa layanan pembayaran ini telah diadopsi secara luas oleh konsumen *e-commerce*, sehingga relevan untuk dianalisis sebagai salah satu indikator perilaku transaksi. Selain itu, tingginya total nilai transaksi mengindikasikan aktivitas ekonomi digital yang signifikan, yang memperkuat kebutuhan akan pendekatan analisis berbasis data untuk memahami heterogenitas perilaku konsumen. Ringkasan ini menjadi dasar awal dalam mengarahkan tahapan analisis eksploratif dan proses segmentasi konsumen selanjutnya.

Secara keseluruhan, rancangan aplikasi dan *layout* sistem yang dikembangkan berfungsi sebagai kerangka utama dalam menyajikan hasil analisis segmentasi secara terstruktur dan mudah dipahami. Desain antarmuka yang sederhana dan informatif memungkinkan pengguna untuk memperoleh gambaran awal mengenai hasil *clustering* sebelum melakukan eksplorasi lebih lanjut. Rancangan ini menjadi dasar bagi tahap implementasi aplikasi dan penyajian *output* analisis yang dibahas pada subbab berikutnya.



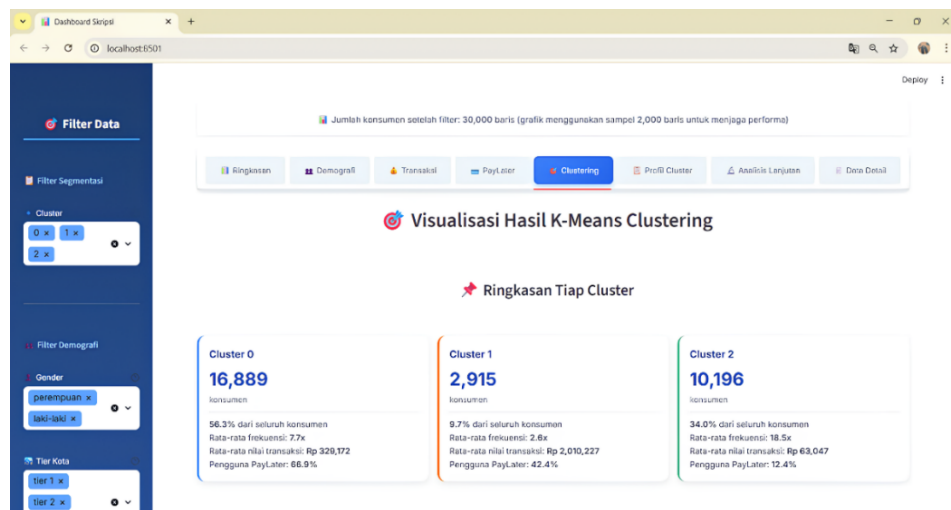
Gambar 2. Tampilan Ringkasan Umum *Dashboard* Interaktif

Implementasi Aplikasi dan Tampilan Output

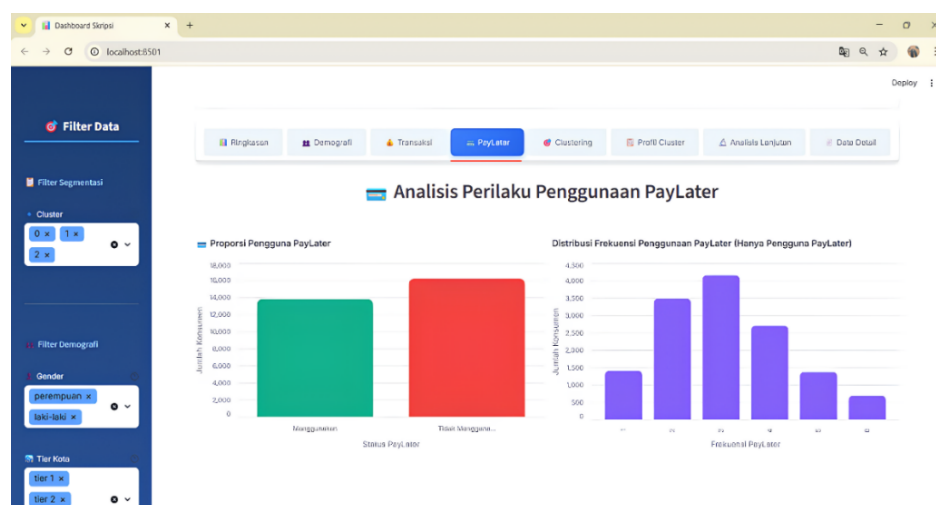
Aplikasi *dashboard* diimplementasikan untuk menampilkan hasil pengolahan data dan segmentasi konsumen secara interaktif. Pengguna dapat melihat ringkasan tiap kluster, proporsi anggota kluster, serta perbandingan karakteristik antar segmen secara dinamis melalui elemen visual yang tersedia.

Selain itu, aplikasi juga menyediakan tampilan khusus untuk menampilkan ringkasan statistik setiap kluster, sehingga memudahkan interpretasi perbedaan perilaku konsumen antar kelompok hasil *clustering*. Gambar 3 menyajikan ringkasan hasil segmentasi konsumen *e-commerce* berdasarkan metode *K-Means Clustering* dengan jumlah kluster optimal sebanyak tiga kluster. Setiap kluster ditampilkan bersama informasi jumlah anggota dan proporsi terhadap keseluruhan populasi konsumen. Perbedaan ukuran kluster menunjukkan bahwa distribusi perilaku konsumen bersifat tidak merata, dengan adanya kelompok dominan serta kelompok dengan karakteristik khusus.

Kluster dengan jumlah anggota terbesar merepresentasikan kelompok konsumen dengan aktivitas belanja menengah dan relatif stabil, sedangkan kluster dengan jumlah anggota lebih kecil mencerminkan konsumen bernilai tinggi dengan nilai transaksi yang besar. Kluster lainnya menunjukkan konsumen dengan intensitas transaksi tinggi namun nilai pembelian relatif rendah. Ringkasan ini menegaskan bahwa pendekatan *clustering* mampu mengelompokkan konsumen ke dalam segmen yang memiliki karakteristik perilaku berbeda secara jelas dan terukur.



Gambar 3. Ringkasan Tiap Kluster Hasil K-Means



Gambar 4. Proporsi Pengguna dan Frekuensi Pemakaian *Paylater*

Gambar 4 memperlihatkan distribusi proporsi konsumen berdasarkan status penggunaan *PayLater* serta frekuensi pemakaiannya. Visualisasi ini menunjukkan perbedaan yang jelas antara konsumen yang menggunakan dan tidak menggunakan layanan *PayLater*, sekaligus mengilustrasikan variasi intensitas penggunaan di antara pengguna *PayLater*.

Sebagian besar pengguna *PayLater* tercatat memiliki frekuensi penggunaan pada tingkat rendah hingga menengah, sementara hanya sebagian kecil konsumen yang menggunakan layanan ini secara intensif. Temuan ini mengindikasikan bahwa meskipun adopsi *PayLater* cukup tinggi, tingkat ketergantungan pengguna terhadap layanan tersebut bervariasi. Pola ini menjadi dasar penting dalam proses segmentasi, karena perbedaan intensitas penggunaan *PayLater* berkontribusi dalam membentuk karakteristik kluster konsumen yang berbeda pada tahap analisis selanjutnya.

Dashboard memungkinkan pengguna untuk memantau distribusi kluster, mengeksplorasi perilaku konsumen, serta membandingkan karakteristik antar segmen secara dinamis. Integrasi ini menjadikan hasil penelitian tidak hanya bersifat analitis, tetapi juga aplikatif sebagai alat bantu pengambilan keputusan berbasis data.

Deskripsi Dataset dan Hasil Analisis Awal

Dataset yang digunakan dalam penelitian ini terdiri dari 30.000 data konsumen *e-commerce* di Indonesia yang mencakup informasi demografis, perilaku transaksi, serta penggunaan layanan *Paylater* selama satu periode tahunan. Dataset diperoleh dari laporan Kredivo dan Katadata Insight Center dan digunakan untuk merepresentasikan perilaku konsumen secara agregat.

Hasil pemeriksaan awal menunjukkan bahwa seluruh variabel terbaca dengan baik dan tidak ditemukan nilai kosong. Kondisi ini menunjukkan bahwa dataset memiliki kualitas yang memadai untuk dianalisis lebih lanjut tanpa memerlukan proses imputasi data. Statistik deskriptif awal juga menunjukkan variasi perilaku konsumen yang cukup tinggi, terutama pada variabel transaksi dan penggunaan *Paylater*.

Hasil deskripsi dataset dan pemeriksaan awal menunjukkan bahwa data yang digunakan memiliki kualitas yang memadai dan struktur yang konsisten untuk dianalisis lebih lanjut. Variasi nilai pada variabel transaksi dan penggunaan *Paylater* mengindikasikan adanya heterogenitas perilaku konsumen, yang menjadi landasan penting bagi analisis eksploratif dan proses segmentasi selanjutnya.

Tabel 1. Deskripsi Variabel Dataset Penelitian

No	Nama Variabel	Keterangan
1.	gender	Jenis kelamin konsumen
2.	age	Usia konsumen
3.	tier_kota	Klasifikasi kota konsumen
4.	status_perkawinan	Status pernikahan
5.	jumlah_anak	Jumlah anak konsumen
6.	frek_transaksi	Jumlah transaksi konsumen
7.	nilai_transaksi	Total nilai belanja tahunan
8.	kategori_produk	Kategori produk utama
9.	paylater_pengguna	Status penggunaan <i>Paylater</i>
10.	frek_paylater	Frekuensi <i>Paylater</i>
11.	kepuasam	Skor Kepuasan Konsumen

Tabel 2. Ringkasan Statistik Deskriptif Dataset

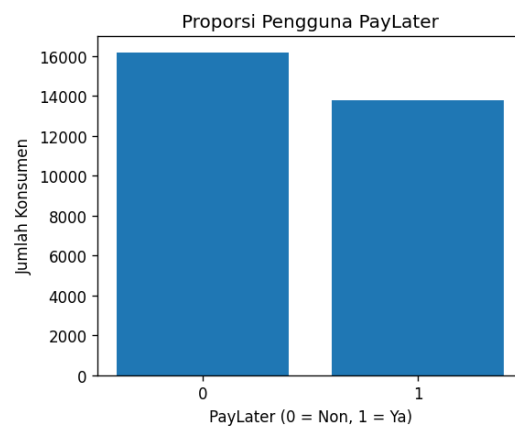
No	Pemeriksaan	Hasil Temuan
1.	Jumlah data	30.000 konsumen
2.	Nilai kosong	Tidak ditemukan
3.	Rentang usisa	18-80 tahun
4.	Rata-rata jumlah anak	1 anak per konsumen
5.	Rata-rata frekuensi transaksi	10,6 transaksi per tahun
6.	Rata-rata nilai transaksi	Rp. 315.718 per konsumen
7.	Proporsi pengguna <i>paylater</i>	46% dari total konsumen
8.	Rata-rata frekuensi penggunaan <i>paylater</i>	1,4 kali per tahun
9.	Skor kepuasan konsumen	72-998 dengan Rata-rata 739

Hasil Analisis Eksploratif Data

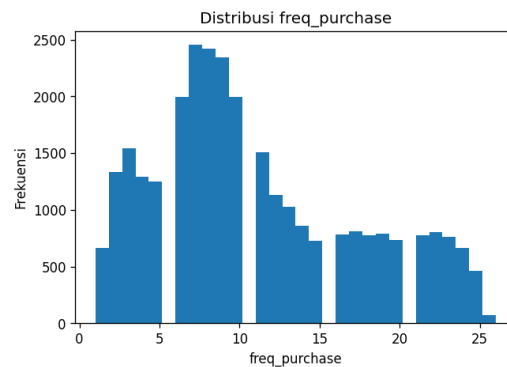
Analisis eksploratif data dilakukan untuk memahami distribusi awal dan karakteristik perilaku konsumen sebelum proses *clustering*. Fokus analisis diarahkan pada variabel transaksi dan penggunaan *Paylater* yang menjadi dasar segmentasi.

Gambar 5 menunjukkan bahwa proporsi konsumen yang menggunakan dan tidak menggunakan *PayLater* relatif seimbang, dengan kecenderungan jumlah pengguna *PayLater* sedikit lebih tinggi. Temuan ini mengindikasikan bahwa *PayLater* telah menjadi metode pembayaran yang cukup umum digunakan, sehingga relevan untuk dijadikan variabel penting dalam analisis perilaku dan segmentasi konsumen.

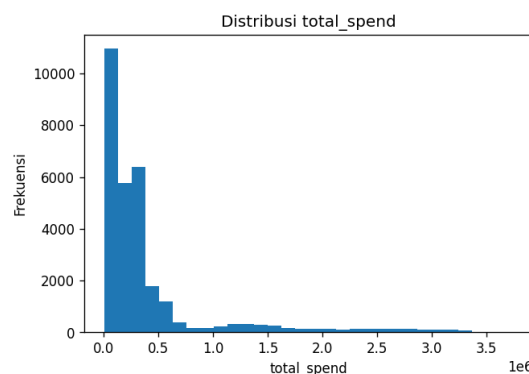
Distribusi pada Gambar 6 menunjukkan bahwa sebagian besar konsumen memiliki frekuensi transaksi pada tingkat rendah hingga menengah, sementara hanya sebagian kecil yang bertransaksi dengan frekuensi tinggi. Pola ini mencerminkan heterogenitas intensitas belanja konsumen, yang menjadi dasar penting dalam pembentukan kluster perilaku.



Gambar 5. Proporsi Pengguna *Paylater*

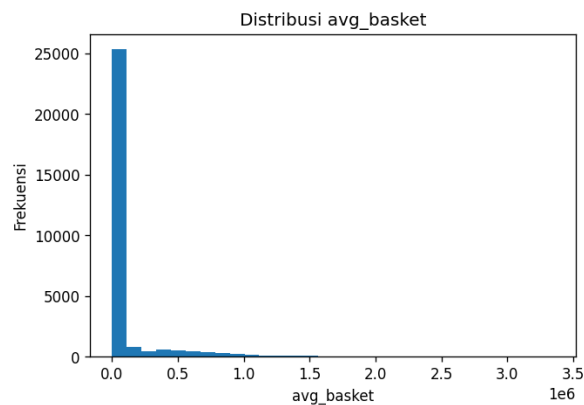


Gambar 6. Distribusi Frekuensi Pembelian

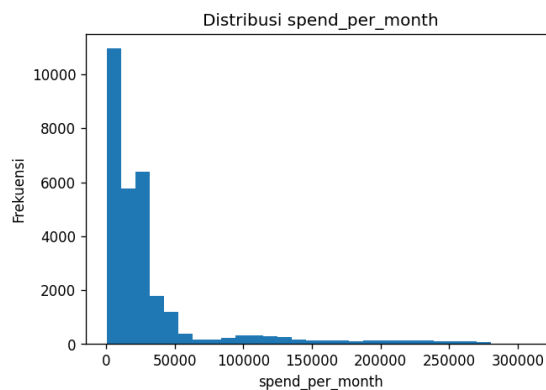


Gambar 7. Distribusi Total Pengeluaran

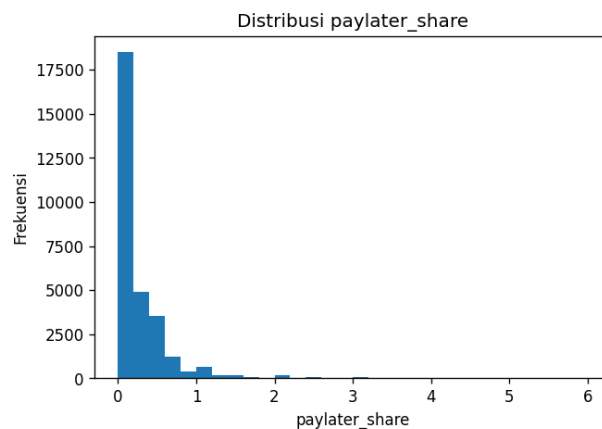
Gambar 7 memperlihatkan distribusi total pengeluaran yang tidak simetris, dengan mayoritas konsumen memiliki nilai pengeluaran relatif rendah dan sebagian kecil konsumen dengan pengeluaran sangat tinggi. Pola ini menunjukkan adanya konsumen bernilai tinggi (*high-value customers*) yang berpotensi membentuk segmen tersendiri dalam proses *clustering*. Distribusi pada Gambar 8 menunjukkan bahwa sebagian besar konsumen memiliki nilai transaksi rata-rata yang rendah, sementara terdapat sejumlah kecil konsumen dengan nilai transaksi yang jauh lebih besar. Ketimpangan ini mengindikasikan perbedaan pola belanja antar konsumen dan memperkuat perlunya proses standarisasi sebelum analisis *clustering*. Gambar 9 menampilkan pola distribusi pengeluaran bulanan yang cenderung miring ke kanan, dengan mayoritas konsumen berada pada tingkat pengeluaran rendah hingga menengah. Variasi nilai yang cukup lebar menunjukkan adanya perbedaan kemampuan dan intensitas belanja antar konsumen.



Gambar 8. Distribusi Rata-Rata Nilai Transaksi



Gambar 9. Distribusi Pengeluaran Bulanan Konsumen



Gambar 10. Distribusi Proporsi Penggunaan Paylater

Gambar 10 menunjukkan bahwa sebagian besar konsumen memiliki proporsi penggunaan *PayLater* yang rendah, sementara hanya sebagian kecil yang menggunakan *PayLater* secara dominan dalam transaksi mereka. Hasil ini mengindikasikan bahwa intensitas penggunaan *PayLater* bervariasi dan berpotensi menjadi pembeda penting dalam segmentasi perilaku konsumen.

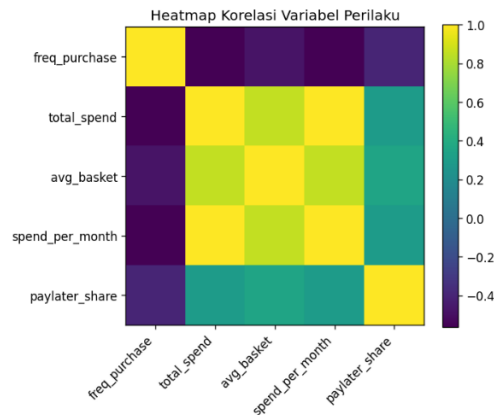
Secara keseluruhan, hasil *EDA* menunjukkan bahwa data memiliki heterogenitas tinggi dan rentang nilai yang berbeda antar variabel. Kondisi ini menguatkan perlunya pendekatan standarisasi dan segmentasi berbasis *clustering*.

Gambar 11 menunjukkan bahwa total pengeluaran, rata-rata nilai transaksi (*average basket*), dan pengeluaran per bulan memiliki korelasi positif yang kuat, menandakan keterkaitan yang erat antar variabel nilai transaksi. Sebaliknya, frekuensi transaksi memiliki korelasi yang relatif rendah terhadap variabel pengeluaran, yang mengindikasikan bahwa tingginya frekuensi belanja tidak selalu diikuti oleh nilai transaksi yang besar. Variabel *paylater share* menunjukkan korelasi sedang terhadap variabel pengeluaran, sehingga berperan sebagai faktor pendukung dalam analisis perilaku konsumen.

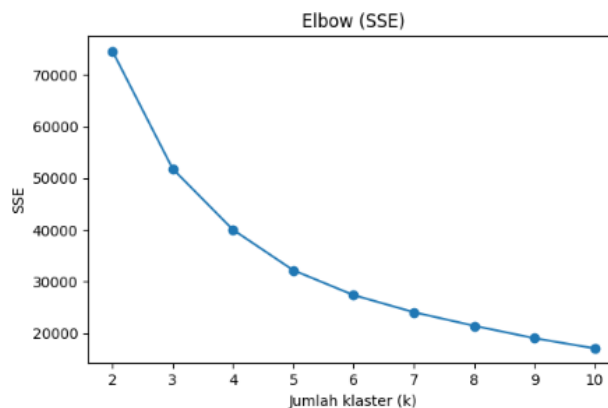
Hasil korelasi menunjukkan hubungan positif yang kuat antara total pengeluaran, rata-rata nilai transaksi, dan pengeluaran bulanan. Sementara itu, frekuensi transaksi tidak selalu berkorelasi langsung dengan besarnya nilai transaksi, yang mengindikasikan kompleksitas perilaku konsumen.

Penentuan Jumlah Kluster Optimal

Penentuan jumlah kluster dilakukan menggunakan pendekatan multi-metrik untuk memperoleh struktur kluster yang seimbang dan mudah diinterpretasikan. Pendekatan ini digunakan untuk menghindari pemilihan jumlah kluster yang hanya bergantung pada satu indikator evaluasi, sehingga hasil segmentasi menjadi lebih objektif. Dengan mempertimbangkan beberapa metrik secara bersamaan, struktur kluster yang dihasilkan diharapkan mampu merepresentasikan perbedaan perilaku konsumen secara lebih jelas.



Gambar 11. *Heatmap* Korelasi Antar Variabel Perilaku

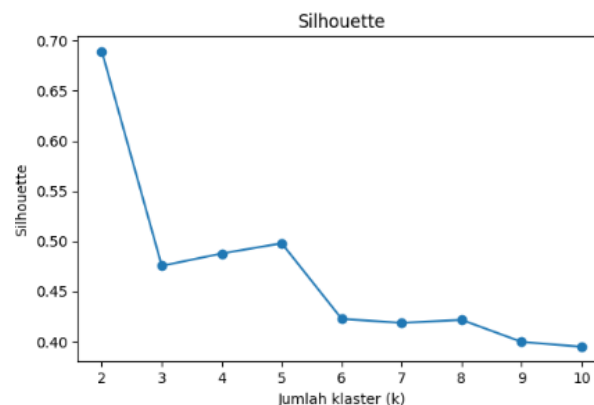


Gambar 12. Grafik *Sum of Squared Errors*

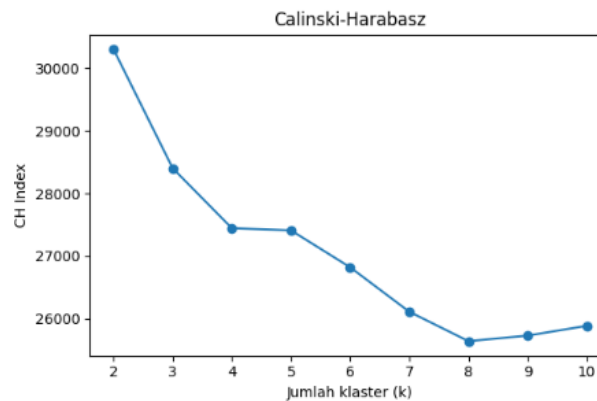
Gambar 12 menunjukkan bahwa Nilai *SSE* mengalami penurunan tajam dari $k = 2$ ke $k = 3$, kemudian menurun secara lebih landai pada jumlah kluster berikutnya. Pola ini menunjukkan adanya titik siku pada $k = 3$, yang mengindikasikan bahwa penambahan kluster setelah nilai tersebut tidak memberikan peningkatan kualitas kluster yang signifikan.

Silhouette Score tertinggi diperoleh pada $k = 2$, yang menunjukkan pemisahan kluster paling optimal secara matematis. Namun, pada $k = 3$ nilai *Silhouette Score* masih berada pada tingkat yang cukup baik dan lebih stabil dibandingkan jumlah kluster yang lebih besar, sehingga tetap layak dipertimbangkan dalam analisis.

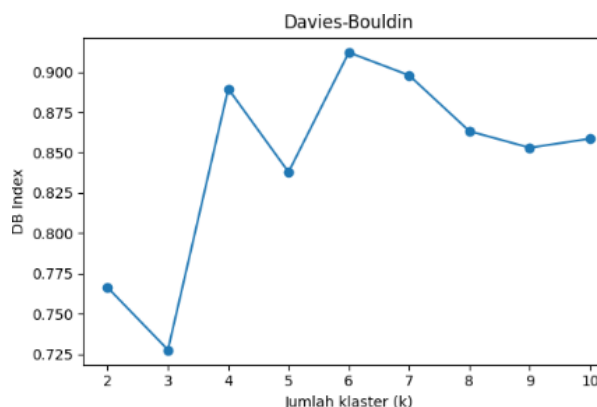
Nilai *Calinski-Harabasz Index* tertinggi diperoleh pada $k = 2$ dan cenderung menurun seiring bertambahnya jumlah kluster. Meskipun demikian, nilai pada $k = 3$ masih relatif tinggi, yang menunjukkan struktur kluster yang cukup baik dengan kompleksitas model yang lebih seimbang.



Gambar 13. Grafik *Silhouette Score*



Gambar 14. Grafik *Calinski Harabasz Index*



Gambar 15. Grafik *Davies Bouldin Index*

Nilai *Davies–Bouldin Index* terendah diperoleh pada $k = 3$, yang menunjukkan tingkat kemiripan antar kluster paling rendah dibandingkan jumlah kluster lainnya. Hal ini mengindikasikan bahwa pada $k = 3$ kluster yang terbentuk memiliki pemisahan yang relatif lebih baik.

Meskipun nilai *Silhouette Score* dan *Calinski–Harabasz Index* tertinggi diperoleh pada $k = 2$, nilai *Davies–Bouldin Index* terendah serta pola titik siku pada grafik *SSE* menunjukkan bahwa $k = 3$ memberikan struktur kluster yang lebih seimbang. Oleh karena itu, penelitian ini menetapkan tiga kluster sebagai hasil akhir segmentasi.

Hasil Segmentasi Konsumen

Penerapan algoritma K-Means dengan jumlah kluster optimal $k = 3$ menghasilkan distribusi konsumen yang berbeda pada masing-masing kluster. Perbedaan ukuran kluster mencerminkan variasi perilaku konsumen *e-commerce* di Indonesia, di mana terdapat kelompok dominan, kelompok bernilai tinggi, dan kelompok dengan aktivitas transaksi tinggi namun nilai pembelian rendah.

Tabel 3. Hasil Evaluasi Jumlah Kluster

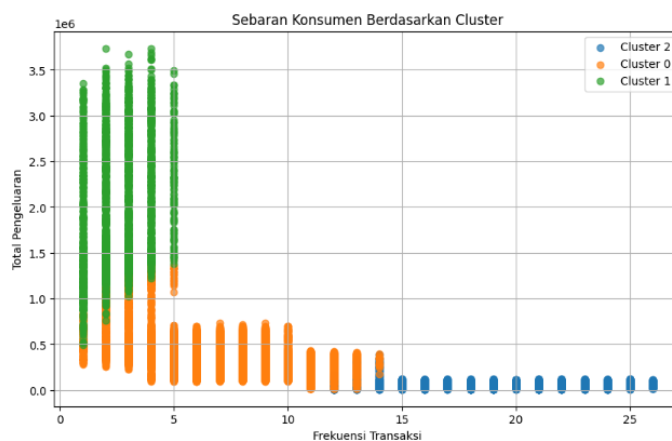
k	SSE	Silhouette	Calinski Harabsz	Davies Bouldin
2	74,6	0.689	30,302	0.766
3	51,836	0.475	28,403	0.728
4	40,053	0.487	27,446	0.889
5	32,220	0.497	27,411	0.837
6	27,416	0.422	26,822	0.912
7	24,101	0.418	26,111	0.897
8	21,475	0.421	25,642	0.863
9	19,075	0.399	25,731	0.853
10	17,106	0.394	25,886	0.858

Tabel 4. Jumlah Anggota pada Setiap Kluster

Cluster	Jumlah Anggota
0	16.889
1	2.915
2	10.196

Tabel 5. Ringkasan Statistik Tiap Kluster

Variabel	Cluster 0	Cluster 1	Cluster 2
Frekuensi transaksi	7.71	2.60	18.51
Total belanja	329,175	2,010,227	63,046
Average basket	62,930	912,853	3,624
Pengeluaran per bulan	27,439	167,518	5,253
Paylater share	0.295	0.632	0.020



Gambar 16. Sebaran Konsumen Berdasarkan Kluster

Gambar 16 memperlihatkan perbedaan pola sebaran konsumen pada masing-masing klaster berdasarkan frekuensi transaksi dan total pengeluaran. Setiap klaster menunjukkan karakteristik yang berbeda, baik dari sisi intensitas transaksi maupun nilai belanja, yang menegaskan bahwa hasil klasterisasi mampu memisahkan konsumen ke dalam segmen perilaku yang jelas dan terukur.

Pembahasan Profil Klaster

Berdasarkan karakteristik yang terbentuk, setiap klaster memiliki ciri khas profil perilaku masing-masing. Klaster pertama merepresentasikan konsumen dengan aktivitas belanja menengah dan penggunaan *Paylater* pada tingkat moderat. Kelompok ini mencerminkan konsumen yang relatif stabil dan konsisten dalam berbelanja.

Klaster kedua merupakan kelompok konsumen bernilai tinggi (*high-value customers*) yang memiliki nilai pembelian per transaksi sangat besar serta kecenderungan penggunaan *Paylater* yang tinggi. Konsumen pada klaster ini berpotensi menjadi target utama strategi retensi dan layanan premium. Klaster ketiga menunjukkan konsumen dengan frekuensi transaksi tinggi namun nilai pembelian rendah. Kelompok ini cenderung sensitif terhadap harga dan lebih sering melakukan pembelian bernilai kecil.

Penelitian ini berbeda dari penelitian sebelumnya karena tidak hanya menerapkan metode K-Means *Clustering* untuk segmentasi konsumen, tetapi juga memanfaatkan variabel perilaku transaksi hasil *feature engineering*, termasuk indikator penggunaan *PayLater*, untuk membentuk profil klaster yang lebih komprehensif. Berbeda dengan studi terdahulu yang umumnya menyajikan hasil *clustering* secara statis, penelitian ini mengintegrasikan hasil segmentasi ke dalam dashboard interaktif berbasis *Streamlit* yang memungkinkan eksplorasi data secara real-time dan mendukung pengambilan keputusan berbasis data. Selain itu, fokus penelitian pada data transaksi aktual konsumen *e-commerce* di Indonesia memberikan konteks analisis yang lebih relevan terhadap karakteristik pasar digital nasional.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan metode K-Means *Clustering* untuk menganalisis dan melakukan segmentasi perilaku konsumen *e-commerce* di Indonesia, serta mengintegrasikan hasil analisis ke dalam sistem *dashboard* interaktif berbasis *Streamlit*. Berdasarkan evaluasi multi-metrik yang meliputi *Sum of Squared Errors*, *Silhouette Score*, *Calinski–Harabasz Index*, dan *Davies–Bouldin Index*, jumlah klaster optimal ditetapkan sebanyak tiga klaster. Hasil segmentasi menunjukkan bahwa perilaku konsumen *e-commerce* bersifat heterogen, baik dari sisi frekuensi transaksi, nilai pembelian, maupun kecenderungan penggunaan layanan *Paylater*.

Tiga segmen yang terbentuk merepresentasikan konsumen dengan aktivitas belanja menengah, konsumen bernilai tinggi dengan transaksi bernilai besar dan tingkat penggunaan *Paylater* yang relatif tinggi, serta konsumen dengan frekuensi transaksi tinggi namun nilai pembelian relatif rendah. Integrasi hasil analisis ke dalam *dashboard* interaktif memungkinkan penyajian informasi secara komprehensif, informatif, dan mudah dipahami, sehingga mendukung proses eksplorasi data dan pengambilan keputusan berbasis data. Secara keseluruhan, penelitian ini memberikan kontribusi akademik dalam penerapan metode *clustering* untuk segmentasi perilaku konsumen, serta kontribusi praktis bagi pelaku *e-commerce* dan penyedia layanan keuangan digital dalam memahami karakteristik konsumen dan potensi nilai ekonomi pada masing-masing segmen.

DAFTAR PUSTAKA

- [1] V. Juita, V. Pujani, R. Rahim, and R. Rahayu, "Dataset on online impulsive buying behavior of buy now pay later users and non-buy now pay later users in Indonesia using the stimulus-organism-response model," *Data Brief*, vol. 54, Jun. 2024, doi: 10.1016/j.dib.2024.110500.
- [2] Mohd. A. Jabbar and Dr. K. Mahabubullah, "Product Recommendation System," *Indian Journal of Computer Science and Technology*, pp. 23–28, Sep. 2025, doi: 10.59256/indjst.20250403005.
- [3] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustainability (Switzerland)*, vol. 14, no. 12, Jun. 2022, doi: 10.3390/su14127243.
- [4] S. A. Ali, A. Razzaq, A. Ali, K. Mohammed, and M. K. Moinuddin, "Customer Segmentation Using Machine Learning," Dec. 2024. [Online]. Available: www.ijfmr.com

- [5] B. I. Nugroho, A. Rafhina, P. S. Ananda, and G. Gunawan, "Journal of Intelligent Decision Support System (IDSS) Customer segmentation in sales transaction data using k-means clustering algorithm," 2024.
- [6] E. Omol, D. Onyangor, L. Mburu, and P. Abuonji, "Application Of K-Means Clustering For Customer Segmentation In Grocery Stores In Kenya," 2024. [Online]. Available: <http://ijstm.inarah.co.id>
- [7] A. Sakinah and D. S. Awaliyah, "Optimization E-Commerce Consumer Segmentation Based On K-Means Clustering And Machine Learning," *Journal of Mathematics, Computations and Statistics*, vol. 8, no. 2, pp. 606–619, Sep. 2025, doi: 10.35580/jmathcos.v8i2.9548.
- [8] S. N. Lathifah and Z. F. Azzahra, "AI-Driven Customers Segmentation Using K-Means Clustering," *G-Tech: Jurnal Teknologi Terapan*, vol. 9, no. 1, pp. 320–329, Jan. 2025, doi: 10.70609/gtech.v9i1.6202.
- [9] F. D. S. M. N. R. A. M. A. A. R. A. P. D. P. P. S. Farrikh Alzami, "Implementation of RFM Method and K-Means Algorithm for Customer Segmentation in E-Commerce with Streamlit," 2023.
- [10] R. Siagian, P. Sirait, and A. Halim, "SISTEMASI: Jurnal Sistem Informasi Penerapan Algoritma K-Means dan K-Medoids untuk Segmentasi Pelanggan pada Data Transaksi E-Commerce The Implementation of K-Means and K-Medoids Algorithm for Customer Segmentation on E-commerce Data Transactions," Feb. 2022. [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [11] H. Ania, M. Mahyuddin, and E. M. Zamzami, "Customer Segmentation of Mobile Banking Users Using Feature Engineering and K-Means Clustering," *Journal La Multiapp*, vol. 6, no. 3, pp. 714–724, Jul. 2025, doi: 10.37899/journallamultiapp.v6i3.2377.
- [12] A. W. Bennett and S. Müller Loose, "User-Centered Development of an Online Dashboard Tool for Economic Sustainability for Small and Medium Enterprises," *Sustainability (Switzerland)*, vol. 16, no. 2, Jan. 2024, doi: 10.3390/su16020557.
- [13] W. Ouyang, "Data Visualization in Big Data Analysis: Applications and Future Trends," *Journal of Computer and Communications*, vol. 12, no. 11, pp. 76–85, 2024, doi: 10.4236/jcc.2024.1211005.
- [14] M. Alves Gomes and T. Meisen, "A review on customer segmentation methods for personalized customer targeting in e-commerce use cases," *Information Systems and e-Business Management*, vol. 21, no. 3, pp. 527–570, Sep. 2023, doi: 10.1007/s10257-023-00640-4.
- [15] Kredivo; Katadata Insight Center, "Laporan Perilaku Konsumen e-Commerce Indonesia 2023," 2023. Accessed: Nov. 07, 2025. [Online]. Available: <https://kredivocorp.com/wp-content/uploads/2023/06/2023-Indonesia-e-Commerce-Behavior-Report.pdf>
- [16] M. Ehsan, "Predictive Modeling for Customer Segmentation in E-Commerce," Jul. 2024, doi: 10.13140/RG.2.2.18926.52805.
- [17] Z. Wu, L. Jin, J. Zhao, L. Jing, and L. Chen, "Research on Segmenting E-Commerce Customer through an Improved K-Medoids Clustering Algorithm," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/9930613.
- [18] T. Caliński and J. Harabasz, "A Dendrite Method For Cluster Analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974, doi: 10.1080/03610927408827101.
- [19] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, 1979, doi: 10.1109/TPAMI.1979.4766909.
- [20] S. Somavarapu and V. Gupta, "Issue 1 www.jetir.org (ISSN-2349-5162)," 2025. [Online]. Available: www.jetir.org