JIKO (JURNAL INFORMATIKA DAN KOMPUTER)
Oktober 2025, Volume: **9**, No. **3** | Pages 560–571
doi: 10.26798/jiko.v9i3.1669
e-ISSN : 2477-3964 – p-ISSN : 2477-4413

ARTICLE

SEGMENTASI NASABAH KARTU KREDIT MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN POLA TRANSAKSI UNTUK PENENTUAN PROFIL NASABAH

CREDIT CARD CUSTOMER SEGMENTATION USING K-MEANS ALGORITHM BASED ON TRANSACTION PATTERNS FOR CUSTOMER PROFILING

Irfan Budiyanto,[*,1] Arief Hermawan,[1] dan Donny Avianto[2]

[1]Program Studi Teknologi Informasi, Program Magister, Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia
[2]Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia
*Penulis Korespondensi: irfan.6240211004@student.uty.ac.id

(Disubmit 03-01-25; Diterima 03-07-25; Dipublikasikan online pada 20-10-25)

Abstrak

Segmentasi nasabah kartu kredit menjadi tantangan penting untuk meningkatkan efektivitas strategi pemasaran dan personalisasi layanan. Permasalahan umum yang dihadapi adalah kurangnya pemahaman mendalam terhadap perilaku transaksi nasabah yang beragam. Secara spesifik, banyak pendekatan segmentasi yang belum optimal dalam menangani distribusi data yang tidak merata dan keberadaan outlier dalam transaksi keuangan. Penelitian ini mengusulkan metode segmentasi berbasis pola transaksi menggunakan algoritma K-Means Clustering dengan pendekatan preprocessing yang diperbarui, yaitu penggunaan RobustScaler untuk mengatasi pengaruh outlier. Seleksi fitur dilakukan secara sistematis melalui backward feature elimination, dengan fokus pada atribut yang merepresentasikan frekuensi dan jenis transaksi. Hasil penelitian menunjukkan bahwa kombinasi fitur optimal menghasilkan tiga cluster utama: nasabah aktif pembelian, nasabah aktif tarik tunai, dan nasabah pasif. Validasi menggunakan Elbow Method dan Silhouette Score menunjukkan konsistensi pada jumlah cluster optimal (k=3), dengan visualisasi PCA memperkuat pemisahan antar segmen. Pendekatan ini mampu menghasilkan segmentasi yang lebih stabil dan representatif, serta dapat digunakan sebagai dasar pengembangan strategi pemasaran yang lebih tertarget. Temuan ini berkontribusi pada peningkatan efisiensi operasional dan kepuasan nasabah melalui layanan yang disesuaikan dengan profil perilaku masing-masing segmen.

Kata kunci: Segmentasi, Kartu Kredit, K-Means Clustering, Elbow Method, Silhouette Score

Abstract

TCustomer segmentation for credit card holders is a crucial challenge for enhancing marketing strategy effectiveness and service personalization. A common issue faced is the lack of a deep understanding of diverse customer transaction behaviors. Specifically, many segmentation approaches are not optimal in handling uneven data distribution and the presence of outliers in financial transactions. This research proposes a transaction pattern-based segmentation method using the K-Means Clustering algorithm with an updated preprocessing approach, namely the use of RobustScaler to address the influence of outliers. Feature selection is performed systematically through backward feature elimination, focusing on attributes representing transaction frequency and type. The research results demonstrate that the optimal feature combination yields three main clusters: active purchasers, active cash withdrawers, and passive customers. Validation using the Elbow Method and

How to Cite: I. Budiyanto *et al.*, "SEGMENTASI NASABAH KARTU KREDIT MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN POLA TRANSAKSI UNTUK PENENTUAN PROFIL NASABAH", *JIKO (JURNAL INFORMATIKA DAN KOMPUTER)*, Volume: **9**, No.**3**, Pages 560–571, Oktober 2025, doi: 10.26798/jiko.v9i3.1669.

Silhouette Score indicates consistency with the optimal number of clusters (k=3), with PCA visualization reinforcing the separation between segments. This approach is capable of generating more stable and representative segmentation, and can be used as a basis for developing more targeted marketing strategies. These findings contribute to increased operational efficiency and customer satisfaction through services tailored to the behavioral profile of each segment.

KeyWords: Segmentation, Credit Card, K Means Clustering, Elbow Method, Silhouette Score

1. Pendahuluan

Kartu kredit kini menjadi alat penting dalam kehidupan modern, tidak hanya sekadar sebagai sarana pembayaran tetapi juga pengelolaan keuangan[1]. Perubahan ini memaksa bank untuk bertransformasi dengan memanfaatkan teknologi terbaru dan analisis data guna memahami profil, perilaku, serta kebutuhan nasabah secara mendalam[2],[3]. Segmentasi pelanggan berbasis data memungkinkan bank menyusun strategi pemasaran lebih personal, efektif dan efisien[4]. Dengan adanya segmentasi nasabah yang jelas, biaya operasional bank khususnya dalam hal pemasaran akan berkurang, karena pemasaran akan lebih personal dan tertarget sehingga tidak ada lagi pemborosan biaya pemasaran[5].

Segmentasi nasabah memungkinkan bank untuk menyesuaikan produk dan layanan mereka dengan kebutuhan spesifik setiap segmen, yang mengarah pada peningkatan kepuasan dan loyalitas nasabah[6]. Misalnya, bank dapat menawarkan produk kartu kredit dengan fitur dan manfaat yang berbeda kepada segmen nasabah tertentu berdasarkan riwayat pengeluaran, preferensi, dan kebutuhan mereka[7]. Selain itu, segmentasi nasabah juga dapat membantu bank dalam berbagai hal seperti meningkatkan efektivitas strategi pemasaran, mengoptimalkan alokasi sumber daya dan memperkuat hubungan dengan nasabah[8],[9]. Oleh karena itu, segmentasi nasabah menjadi strategi yang krusial dalam industri perbankan modern[10].

Terdapat beberapa penelitian yang menggunakan dataset sama yang digunakan dalam penelitian ini. Penelitian yang dilakukan oleh Dwididanti S[11] membandingkan algoritma *Bisecting K-Means* dan *Fuzzy C-Means* dalam segmentasi data pengguna kartu kredit. Dataset yang digunakan berasal dari Kaggle, mencakup 8.951 data pengguna dengan 18 atribut. Setelah melalui *preprocessing* (normalisasi Min-Max dan reduksi dimensi PCA), kedua algoritma diuji menggunakan metrik *silhouette coefficient*. Hasilnya, Bisecting K-Means tanpa normalisasi menunjukkan performa terbaik dengan nilai *silhouette* 0,588, lebih tinggi dibandingkan *Fuzzy C-Means* (0,488 tanpa normalisasi dan 0,582 dengan normalisasi). Kesimpulannya, *Bisecting K-Means* lebih unggul dalam ketepatan pengelompokan data, meskipun normalisasi meningkatkan performa *Fuzzy C-Means*. Penelitian yang dilakukan oleh Ramadhan H[12] membandingkan metode *Hierarchical*, *K-Means*, dan DBSCAN untuk segmentasi pelanggan kartu kredit berdasarkan tingkat pengeluaran. Menggunakan dataset dari Kaggle dan evaluasi dengan *silhouette coefficient*, hasil terbaik diperoleh dari metode *Hierarchical Clustering* dengan dua *cluster*, mencapai skor 0.82322. Temuan ini menunjukkan bahwa pendekatan hierarki paling efektif untuk segmentasi pelanggan, dan disarankan strategi pemasaran dikembangkan untuk dua segmen tersebut. Penelitian yang dilakukan oleh Karo I[13] menggunakan algoritma K-Means untuk melakukan segmentasi pelanggan kartu kredit berdasarkan perilaku penggunaan mereka. Dengan data dari Kaggle berisi 8.950 pengguna dan 18 variabel perilaku, proses *clustering* menghasilkan enam segmen utama: hobi belanja, pembayaran tepat waktu, pembayaran cicilan, penarikan tunai, pembelian barang mewah, dan pengguna pasif. Evaluasi menggunakan indeks *Silhouette* menunjukkan bahwa enam *cluster* adalah jumlah optimal untuk segmentasi yang efektif. Penelitian yang dilakukan oleh Alhamdani F [14] menggunakan metode *K-Means Clustering* untuk segmentasi pelanggan kartu kredit berdasarkan perilaku penggunaan. Dibandingkan dengan *Agglomerative Clustering*, GMM, dan DBSCAN, K-Means menunjukkan performa terbaik dengan nilai *silhouette coefficient* sebesar 0.207014 dan menghasilkan 3 *cluster*. Visualisasi menggunakan T-SNE menunjukkan tiga tipe pengguna: penggunaan moderat, penggunaan rendah, dan penggunaan tinggi. Hasil ini dapat digunakan untuk merancang strategi pemasaran yang lebih efektif.

Penelitian ini merupakan pengembangan dari studi sebelumnya yang dilakukan oleh Alhamdani F[14] yang menunjukkan bahwa metode K-Means memberikan hasil clustering terbaik dibandingkan metode lain seperti DBSCAN, GMM, dan *Agglomerative Clustering*[15]. Namun masih ditemukan kelemahan, yaitu penggunaan metode standarisasi *StandardScaler* pada dataset dengan atribut uyang memiliki jumlah

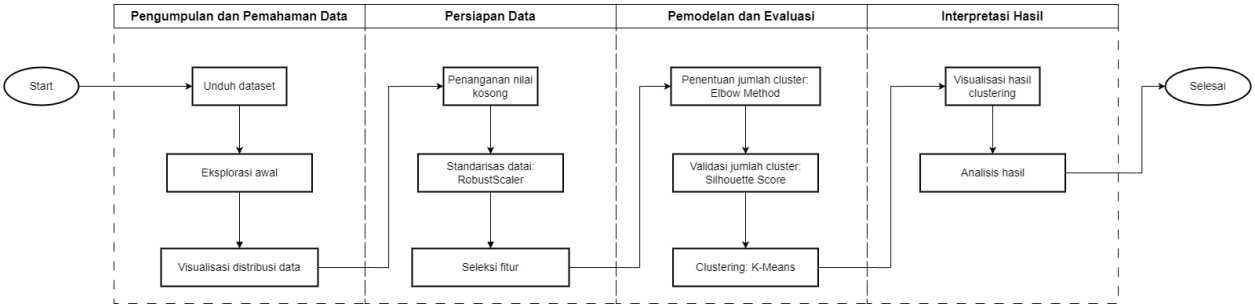
outlier cukup besar. Keberadaan outlier ini berpotensi mengganggu distribusi data dan memengaruhi akurasi hasil *clustering*[16]. Sebagai bentuk peningkatan kualitas, penelitian ini menawarkan pendekatan baru dengan menggunakan scaler adaptif yang lebih tahan terhadap pengaruh *outlier*, sehingga diharapkan dapat menghasilkan segmentasi yang lebih stabil dan representatif[17].

Kebaruan penelitian ini terletak pada penerapan *scaler* adaptif dan eksplorasi kombinasi fitur yang berfokus pada pola transaksi nasabah. Seleksi fitur dilakukan secara sistematis dengan mempertimbangkan korelasi antar atribut dan relevansi langsung terhadap karakteristik perilaku transaksi, menghasilkan kombinasi optimal untuk membentuk tiga *cluster* yang bermakna.

Meski demikian, penelitian ini memiliki beberapa keterbatasan. Pertama, analisis hanya difokuskan pada atribut transaksi tanpa mempertimbangkan faktor demografis atau riwayat kredit yang juga dapat memengaruhi perilaku nasabah. Kedua, validasi hasil *clustering* masih terbatas pada metode internal seperti *silhouette score*, tanpa melibatkan evaluasi eksternal atau uji implementasi strategi pemasaran berbasis *cluster*. Ketiga, penggunaan dataset publik dari Kaggle meskipun bermanfaat, mungkin tidak sepenuhnya mencerminkan kondisi aktual nasabah di institusi keuangan tertentu. Oleh karena itu, penelitian lanjutan disarankan untuk mengintegrasikan data yang lebih sesuai dan memperluas pendekatan validasi agar hasil segmentasi dapat diterapkan secara lebih praktis dalam dunia industri.

2. Metode

Penelitian ini menerapkan metode *data mining* untuk memahami pola-pola tersembunyi dalam data. Prosesnya meliputi langkah-langkah pengumpulan data, *preprocessing* data, teknik clustering, dan analisis mendalam atas hasil *clustering*. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. *Flowchart* Alur Penelitian

Seluruh tahapan dilakukan menggunakan google colab, dan source code lengkap tersedia di github: https://github.com/irfanbudianto/Notebook_Segmentasi_Nasabah_KMeans.git.

2.1 Pengumpulan dan Pemahaman Data

Penelitian ini menggunakan dataset publik berjudul "Bank Customer Segmentation" yang tersedia di platform kaggle (https://www.kaggle.com/datasets/sonugupta0932/bank-customer-segmentation/data). Dataset terdiri dari 8950 entri nasabah dengan 18 atribut yang merepresentasikan berbagai aspek perilaku transaksi kartu kredit. Tabel 1 merangkum deskripsi masing-masing atribut yang digunakan dalam penelitian ini.

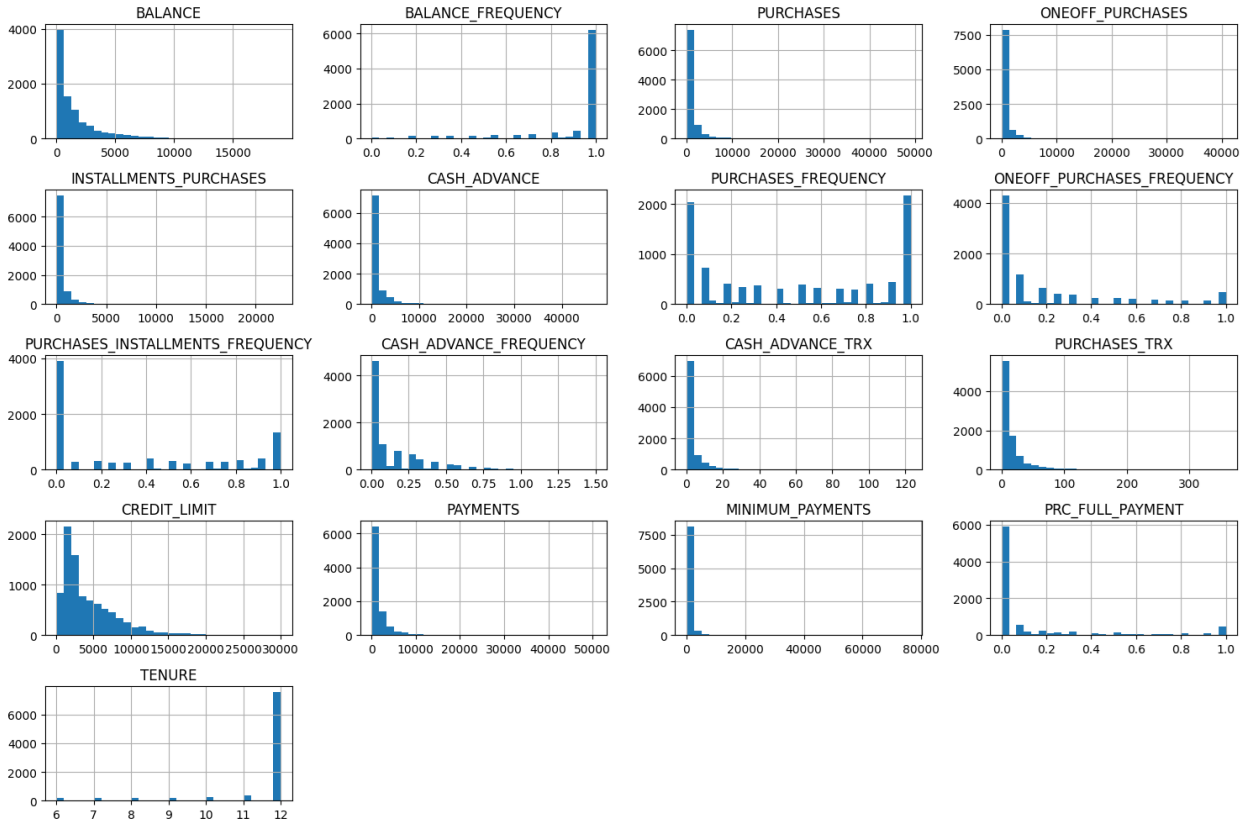
Langkah awal dalam memahami struktur data dilakukan melalui eksplorasi awal menggunakan fungsi df.info() dan df.describe(). Fungsi df.info() memberikan informasi mengenai jumlah entri, tipe data, serta keberadaan nilai kosong pada setiap atribut. Sementara itu, df.describe() menyajikan statistik deskriptif seperti nilai minimum, maksimum, rata-rata, dan kuartil, yang berguna untuk mengidentifikasi distribusi data dan potensi nilai ekstrem[17].

Untuk memperdalam pemahaman terhadap karakteristik data, dilakukan visualisasi awal menggunakan histogram dan boxplot. Histogram pada Gambar 2 menunjukkan bahwa sebagian besar atribut memiliki distribusi yang tidak simetris (*right-skewed*), yang umum ditemukan dalam data transaksi keuangan[18].

Tabel 1. Keterangan Atribut Dataset

Atribut	Deskripsi
CUST_ID	No. Nasabah
BALANCE	Total saldo rekening nasabah
BALANCE_FREQUENCY	Frekuensi update saldo
PURCHASES	Jumlah total pembelian yang dilakukan oleh nasabah
ONEOFF_PURCHASES	Jumlah pembelian untuk satu kali pembayaran
INSTALLMENTS_PURCHASES	Jumlah pembelian yang dibayar dengan cicilan
CASH_ADVANCE	Total tarik tunai yang diambil oleh nasabah
PURCHASES_FREQUENCY	Frekuensi pembelian
ONEOFF_PURCHASES_FREQUENCY	Frekuensi pembelian satu kali
PURCHASES_INSTALLMENTS_FREQUENCY	Frekuensi pembelian cicilan
CASH_ADVANCE_FREQUENCY	Frekuensi penarikan tunai
CASH_ADVANCE_TRX	Jumlah transaksi untuk penarikan tunai
PURCHASES_TRX	Jumlah transaksi pembelian
CREDIT_LIMIT	Batas kredit nasabah
PAYMENTS	Total pembayaran yang dilakukan oleh nasabah
MINIMUM_PAYMENTS	Pembayaran minimum yang dilakukan oleh nasabah
PRC_FULL_PAYMENT	Persentase pembayaran penuh yang dilakukan oleh nasabah
TENURE	Lama menjadi nasabah

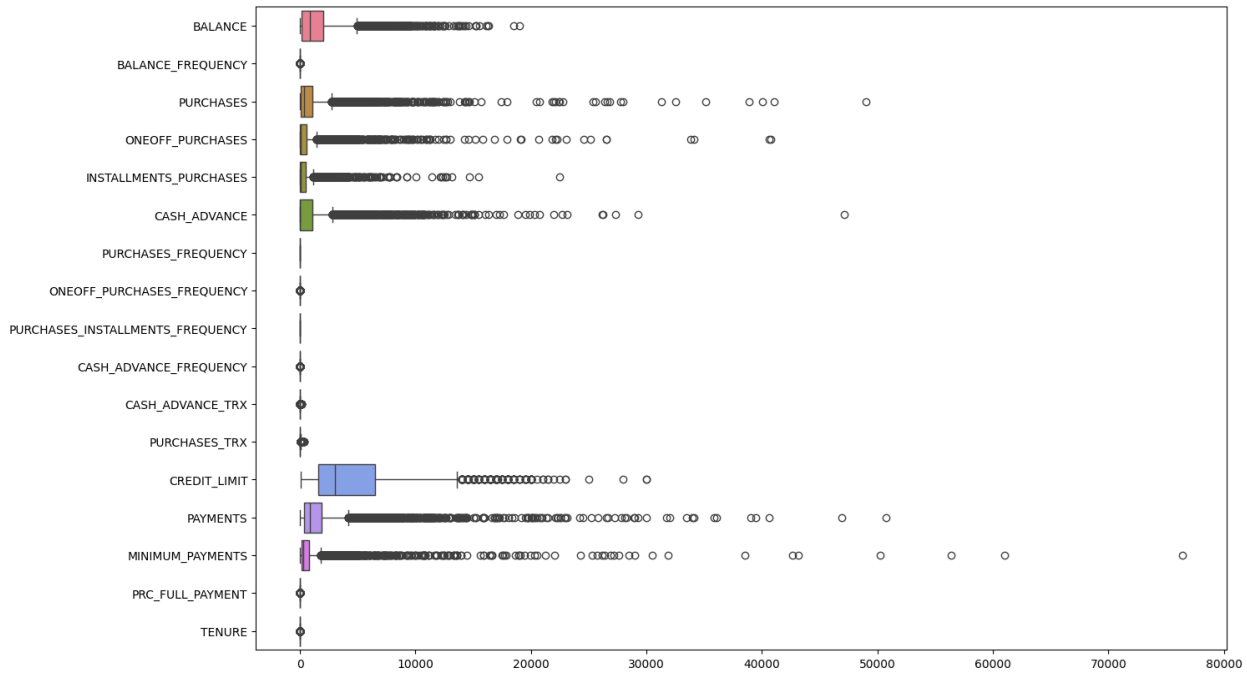
Hal ini mengindikasikan bahwa sebagian besar nasabah memiliki aktivitas transaksi rendah, sementara sebagian kecil menunjukkan aktivitas yang sangat tinggi.



Gambar 2. Histogram Atribut Numerik

Selanjutnya, visualisasi boxplot pada Gambar 3 digunakan untuk mengidentifikasi keberadaan outlier. Beberapa atribut menunjukkan adanya nilai ekstrem yang signifikan, yang ditandai dengan titik-titik di

luar rentang interkuartil. Temuan ini menjadi dasar penting dalam pemilihan metode standarisasi yang lebih tahan terhadap outlier pada tahap preprocessing data[19].



Gambar 3. Boxplot Atribut Numerik

2.2 Persiapan Data

2.2.1 Penanganan Nilai Kosong

Dataset publik yang diperoleh dari sumber seperti kaggle seringkali memiliki kualitas data yang kurang baik. Masalah-masalah yang umum ditemukan antara lain nilai yang hilang, *outlier*, format data yang tidak konsisten, dan data bias[20]. Untuk mengatasi hal ini, diperlukan tahapan persiapan data yang efektif untuk meningkatkan kualitas data sebelum dilanjutkan ke proses selanjutnya[21]. Langkah awal dalam persiapan data ini adalah mengidentifikasi dan mengatasi nilai kosong dalam dataset, sehingga data menjadi lebih lengkap dan siap untuk diproses lebih lanjut. Tabel 2 menunjukkan daftar nilai kosong yang ditemukan dalam seluruh atribut numerik.

Terhadap nilai kosong yang ditemukan pada atribut MINIMUM_PAYMENTS dan CREDIT_LIMIT, dilakukan penanganan dengan mengisi nilai kosong dengan median dari semua nilai tidak kosong dalam atribut tersebut.

2.2.2 Standarisasi Data

Dalam proses eksplorasi data, standarisasi merupakan langkah penting yang bertujuan untuk menyamakan skala antar fitur agar tidak ada atribut yang mendominasi proses analisis hanya karena memiliki nilai numerik yang lebih besar[22]. Standarisasi dilakukan dengan mengubah distribusi data sehingga memiliki rata-rata nol dan deviasi standar satu, atau dengan pendekatan lain yang menyesuaikan skala data tanpa terpengaruh oleh nilai ekstrem. Tujuan utama dari standarisasi adalah untuk memastikan bahwa setiap atribut berkontribusi secara seimbang dalam proses pemodelan, terutama pada algoritma yang sensitif terhadap jarak seperti K-Means. Dalam konteks penelitian ini, setelah dilakukan eksplorasi awal ditemukan adanya *outlier* yang cukup signifikan pada beberapa atribut transaksi. Oleh karena itu, dibanding menggunakan metode standarisasi konvensional seperti StandardScaler yang sangat dipengaruhi oleh *outlier*, penelitian ini memilih menggunakan *RobustScaler*. *RobustScaler* melakukan transformasi data berdasarkan median dan *interkuartil range* (IQR), sehingga lebih tahan terhadap pengaruh nilai-nilai ekstrem[23]. Keunggulan utama RobustScaler dibandingkan scaler lain adalah kemampuannya menjaga kestabilan distribusi data meskipun terdapat *outlier*, sehingga sangat sesuai untuk dataset transaksi keuangan yang secara alami memiliki sebaran nilai yang tidak merata.

Tabel 2. Hasil Pengecekan Nilai Kosong

Atribut	Nilai Kosong
BALANCE	0
BALANCE_FREQUENCY	0
PURCHASES	0
ONEOFF_PURCHASES	0
INSTALLMENTS_PURCHASES	0
CASH_ADVANCE	0
PURCHASES_FREQUENCY	0
ONEOFF_PURCHASES_FREQUENCY	0
PURCHASES_INSTALLMENTS_FREQUENCY	0
CASH_ADVANCE_FREQUENCY	0
CASH_ADVANCE_TRX	0
PURCHASES_TRX	0
CREDIT_LIMIT	1
PAYMENTS	0
MINIMUM_PAYMENTS	313
PRC_FULL_PAYMENT	0
TENURE	0

2.2.3 Seleksi Fitur

Proses seleksi fitur dalam penelitian ini memegang peranan krusial untuk memastikan bahwa segmentasi yang dihasilkan benar-benar merefleksikan pola transaksi nasabah. Pendekatan yang dilakukan tidak hanya berhenti pada analisis awal, melainkan melalui sebuah proses yang sistematis dan iteratif. Tahap pertama adalah seleksi berdasarkan *domain knowledge*. Dari keseluruhan atribut yang tersedia, diidentifikasi sepuluh atribut awal yang secara langsung menggambarkan perilaku transaksional nasabah, seperti tampak dalam Tabel 3. Atribut-atribut ini mencakup berbagai dimensi seperti volume, frekuensi, dan jenis transaksi, baik untuk pembelian maupun penarikan tunai. Tujuan dari tahap ini adalah untuk menyaring dan memfokuskan analisis hanya pada fitur-fitur yang paling relevan dengan tujuan penelitian.

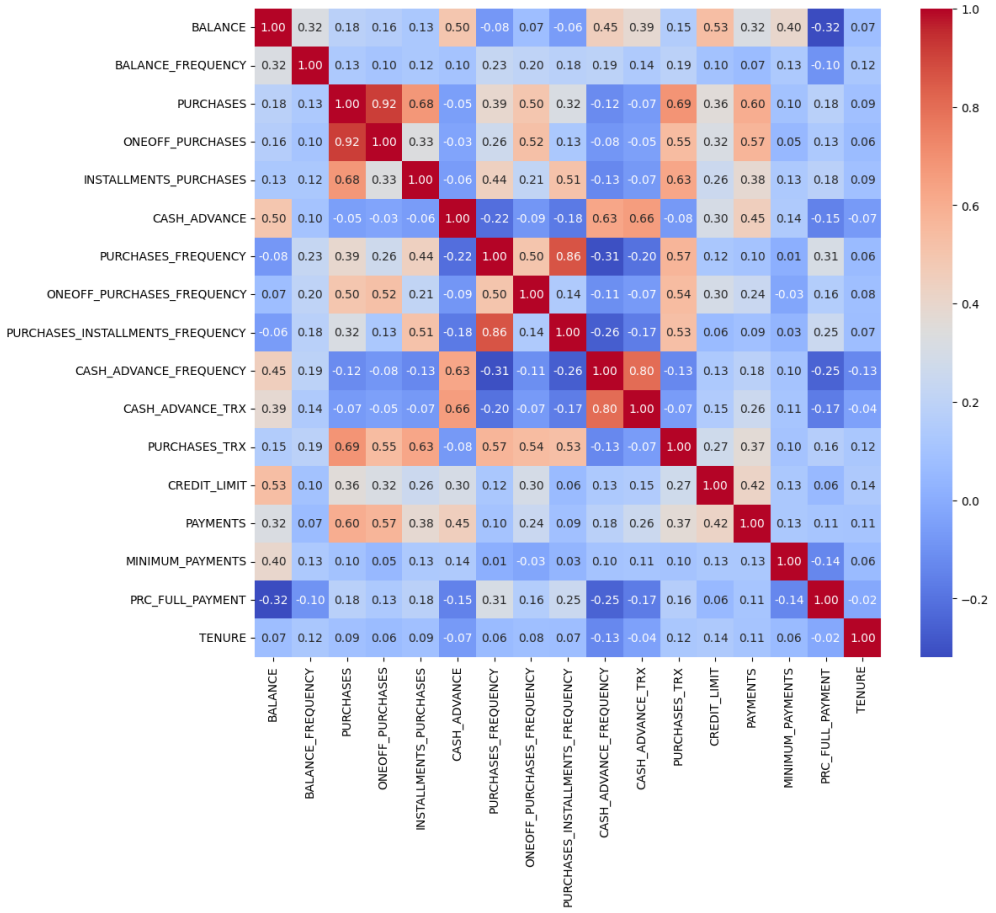
Tabel 3. Atribut Terkait Pola Transaksi

Atribut	Deskripsi
PURCHASES	Jumlah total pembelian yang dilakukan oleh nasabah
ONEOFF_PURCHASES	Jumlah pembelian untuk satu kali pembayaran
INSTALLMENTS_PURCHASES	Jumlah pembelian yang dibayar dengan cicilan
CASH_ADVANCE	Total tarik tunai yang diambil oleh nasabah
PURCHASES_FREQUENCY	Frekuensi pembelian
ONEOFF_PURCHASES_FREQUENCY	Frekuensi pembelian satu kali
PURCHASES_INSTALLMENTS_FREQUENCY	Frekuensi pembelian cicilan
CASH_ADVANCE_FREQUENCY	Frekuensi penarikan tunai
CASH_ADVANCE_TRX	Jumlah transaksi untuk penarikan tunai
PURCHASES_TRX	Jumlah transaksi pembelian

Setelah set atribut awal ditentukan, penelitian ini menerapkan metode seleksi fitur iteratif, yaitu *backward feature elimination*. Metode ini bekerja dengan cara menghilangkan fitur satu per satu dari set fitur awal, lalu pada setiap iterasi dilakukan pengujian ulang untuk menemukan nilai k = 3 menggunakan *elbow method* dan *silhouette score*.

Sebagai alat bantu analisis selama proses iteratif, digunakan matriks korelasi, seperti tampak pada Gambar 4. Matriks ini berfungsi untuk memberikan pemahaman mengenai hubungan linear antar fitur yang sedang diuji. Dengan mengamati korelasi, peneliti dapat membuat keputusan yang lebih informatif saat

mengeliminasi fitur, terutama jika ditemukan adanya redundansi atau multikolinearitas yang tinggi antar fitur yang dapat mengganggu kinerja algoritma K-Means.



Gambar 4. Matrik Korelasi

2.3 Pemodelan dan Evaluasi

2.3.1 Penentuan Jumlah Cluster

Sebelum menerapkan algoritma K-Means untuk proses *clustering*, langkah penting yang harus dilakukan adalah menentukan jumlah *cluster* optimal, atau nilai k. Penentuan nilai k sangat krusial karena akan memengaruhi struktur dan kualitas hasil segmentasi. Penelitian ini merupakan pengembangan dari studi sebelumnya, nilai k yang digunakan tetap sama, yaitu sebanyak tiga *cluster*. Namun, mengingat adanya perubahan pendekatan dalam preprocessing yakni penggunaan RobustScaler yang lebih tahan terhadap *outlier*, diperlukan penyesuaian dalam pemilihan kombinasi fitur agar tetap menghasilkan tiga *cluster* yang bermakna.

Untuk memastikan bahwa jumlah *cluster* yang dipilih sudah tepat, digunakan metode Elbow, yaitu dengan menghitung nilai inersia (jumlah kuadrat jarak antara data dan pusat *cluster*) untuk berbagai nilai k. Grafik inersia kemudian dianalisis untuk menemukan titik "siku" (*elbow point*), yaitu titik di mana penurunan inersia mulai melambat secara signifikan. Titik ini dianggap sebagai jumlah *cluster optimal*. Setelah itu, validasi dilakukan menggunakan *Silhouette Score*, yang mengukur seberapa baik setiap data cocok dengan cluster-nya sendiri dibandingkan dengan cluster lain. Nilai *Silhouette Score* tertinggi yang diperoleh pada k = 3 memperkuat bahwa jumlah *cluster* tersebut adalah yang paling representatif untuk dataset ini.

2.3.2 Clustering K-Means

Setelah jumlah *cluster* optimal ditentukan melalui metode Elbow dan divalidasi dengan *Silhouette Score*, langkah selanjutnya adalah menerapkan algoritma K-Means untuk melakukan proses *clustering*. K-Means merupakan algoritma *unsupervised learning* yang bekerja dengan cara mengelompokkan data ke dalam sejumlah cluster berdasarkan kemiripan fitur. Algoritma ini memulai proses dengan menentukan pusat

cluster secara acak, kemudian secara iteratif memperbarui posisi pusat cluster berdasarkan rata-rata posisi data dalam masing-masing kelompok hingga konvergen. Kelebihan utama K-Means adalah kesederhanaannya, efisiensi komputasi, dan kemampuannya dalam menangani dataset berukuran besar. Pemilihan K-Means dalam penelitian ini didasarkan pada kesinambungan dengan penelitian sebelumnya, yang telah menunjukkan bahwa algoritma ini memberikan hasil segmentasi yang cukup baik untuk data transaksi kartu kredit. Setelah proses *clustering* selesai, hasil pengelompokan ditambahkan ke dalam dataset asli dalam bentuk label baru dengan nama atribut 'Cluster', yang menunjukkan keanggotaan masing-masing nasabah dalam salah satu dari tiga segmen yang terbentuk. Label ini kemudian digunakan dalam tahap analisis lanjutan untuk memahami karakteristik masing-masing *cluster*.

2.4 Interpretasi Hasil

Analisis data akan dilakukan melalui pengamatan hasil visualisasi *cluster* yang dihasilkan oleh algoritma K-Means, dengan menggunakan atribut-atribut yang telah ditentukan sebelumnya. Dalam proses visualisasi ini, atribut-atribut dominan dan tidak dominan dalam tiap *cluster* akan teridentifikasi, sehingga dapat digunakan sebagai representasi dari masing-masing *cluster*. Berdasarkan atribut-atribut ini, *cluster* dapat diberi label yang sesuai dan bermakna, yang kemudian dapat dijadikan acuan dalam strategi mengelola nasabah.

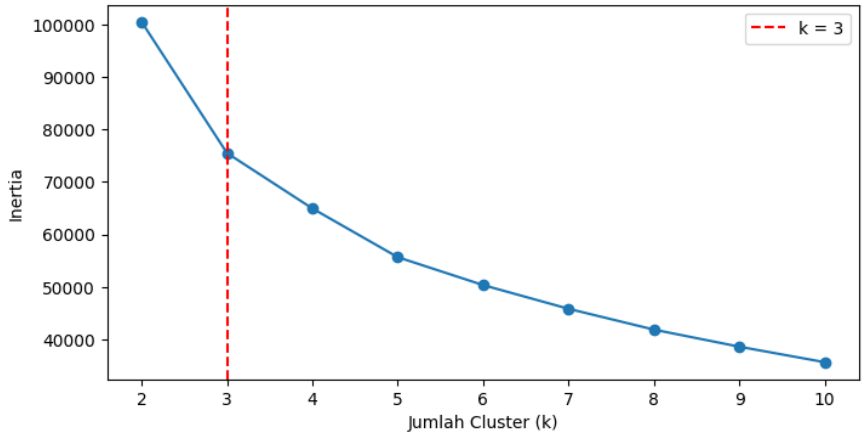
3. Hasil

Hasil proses seleksi fitur secara iteratif menggunakan metode *backward feature elimination* disajikan pada Tabel 4. Proses ini diawali dengan sepuluh fitur awal yang dianggap relevan untuk menghitung nilai k. Apabila hasil evaluasi belum menunjukkan konsistensi antara Elbow Method dan Silhouette Score pada k = 3, maka dilakukan eliminasi fitur secara bertahap, dimulai dari fitur pertama dalam daftar. Proses ini diulang hingga diperoleh kombinasi fitur yang menghasilkan nilai k = 3 secara konsisten pada kedua metode evaluasi tersebut. Setelah melalui tiga iterasi, ditemukan kombinasi fitur optimal yang menunjukkan konsistensi nilai k = 3 baik pada Elbow Method maupun *Silhouette Score*.

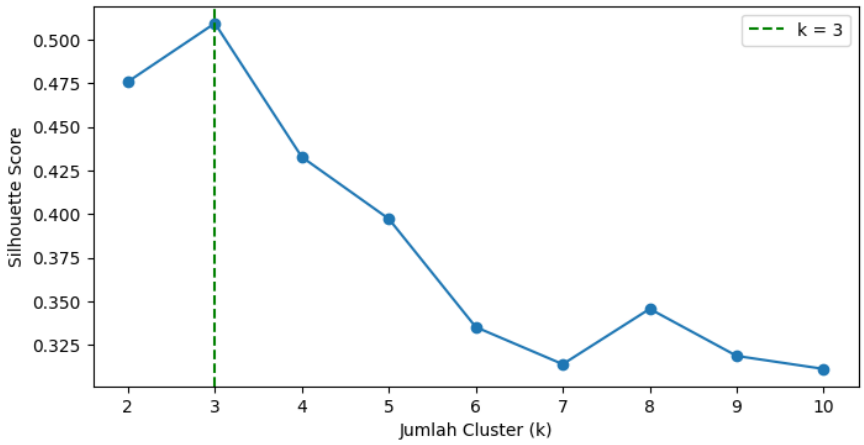
Tabel 4. Hasil Iterasi Pencarian Kombinasi Fitur

Iterasi	Daftar Fitur	k	
		Elbow Method	Silhouette Score
Iterasi-1	'PURCHASES', 'ONEOFF_PURCHASES', 'INSTALLMENTS_PURCHASES', 'CASH_ADVANCE', 'PURCHASES_FREQUENCY', 'ONEOFF_PURCHASES_FREQUENCY', 'PURCHASES_INSTALLMENTS_FREQUENCY', 'CASH_ADVANCE_FREQUENCY', 'CASH_ADVANCE_TRX', 'PURCHASES_TRX'	3	2
Iterasi-2	'ONEOFF_PURCHASES', 'INSTALLMENTS_PURCHASES', 'CASH_ADVANCE', 'PURCHASES_FREQUENCY', 'ONEOFF_PURCHASES_FREQUENCY', 'PURCHASES_INSTALLMENTS_FREQUENCY', 'CASH_ADVANCE_FREQUENCY', 'CASH_ADVANCE_TRX', 'PURCHASES_TRX'	3	2
Iterasi-3	'INSTALLMENTS_PURCHASES', 'CASH_ADVANCE', 'PURCHASES_FREQUENCY', 'ONEOFF_PURCHASES_FREQUENCY', 'PURCHASES_INSTALLMENTS_FREQUENCY', 'CASH_ADVANCE_FREQUENCY', 'CASH_ADVANCE_TRX', 'PURCHASES_TRX'	3	3

Setelah kombinasi fitur optimal ditemukan, dilakukan proses clustering menggunakan algoritma K-Means. Hasil visualisasi Elbow Method pada Gambar 5 menunjukkan titik siku yang jelas pada nilai k = 3, yang mengindikasikan bahwa penambahan cluster lebih lanjut tidak memberikan pengurangan inersia yang signifikan. Selanjutnya, validasi dilakukan menggunakan Silhouette Score, yang juga menunjukkan nilai tertinggi pada k = 3, seperti terlihat pada Gambar 6. Konsistensi antara hasil Elbow dan Silhouette Score memperkuat keyakinan bahwa tiga cluster merupakan jumlah yang paling representatif untuk dataset ini, terutama dalam konteks pola transaksi nasabah.



Gambar 5. Elbow Method



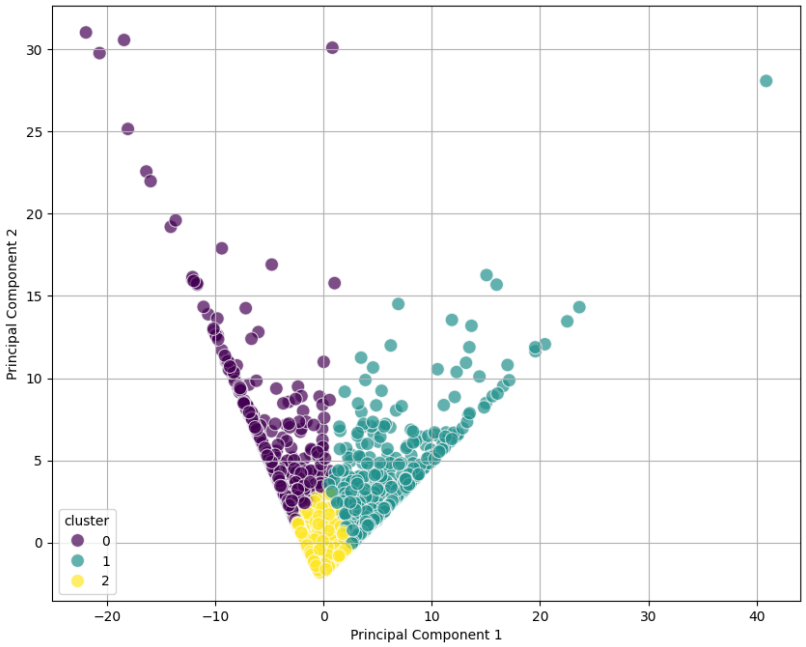
Gambar 6. Silhouette Score

Untuk memperjelas pemisahan antar cluster, dilakukan visualisasi hasil clustering menggunakan metode *Principal Component Analysis* (PCA) dalam dua dimensi. Gambar 7 menunjukkan bahwa ketiga *cluster* yang terbentuk memiliki batas yang cukup jelas dan tidak saling tumpang tindih, menandakan bahwa fitur-fitur yang digunakan berhasil membedakan karakteristik nasabah secara efektif. Visualisasi ini juga membantu dalam interpretasi awal terhadap masing-masing kelompok, sebelum dilakukan analisis lebih lanjut.

Hasil clustering kemudian dikembalikan ke dataset asli dengan menambahkan label baru bernama 'Cluster', yang menunjukkan keanggotaan masing-masing nasabah dalam salah satu dari tiga segmen. Dengan adanya label ini, analisis akhir dapat dilakukan untuk memahami karakteristik dominan dari setiap *cluster*, seperti frekuensi transaksi, preferensi tarik tunai, dan kemampuan membayar. Analisis ini menjadi dasar dalam penentuan strategi pemasaran yang lebih terarah dan sesuai dengan profil masing-masing segmen nasabah.

4. Pembahasan

Hasil segmentasi nasabah berdasarkan pola transaksi menunjukkan pembagian yang cukup jelas ke dalam tiga cluster yang merepresentasikan karakteristik perilaku yang berbeda. Analisis dilakukan terhadap nilai



Gambar 7. *Principal Component Analysis* Hasil Clustering

rata-rata dari delapan atribut utama yang berkaitan dengan aktivitas pembelian dan penggunaan fasilitas tarik tunai. Tabel 5 memperlihatkan hasil segmentasi dengan nilai high (merah), medium, low (hijau) untuk masing-masing atribut.

Tabel 5. Hasil Clustering

Atribut	0	1	2
INSTALLMENTS_PURCHASES	2379.6005	180.3974	241.9765
CASH_ADVANCE	386.3245	4946.6913	419.0329
PURCHASES_FREQUENCY	0.9654	0.2814	0.4736
ONEOFF_PURCHASES_FREQUENCY	0.5564	0.1412	0.1752
PURCHASES_INSTALLMENTS_FREQUENCY	0.8923	0.1802	0.3383
CASH_ADVANCE_FREQUENCY	0.0557	0.5030	0.0858
CASH_ADVANCE_TRX	1.2757	15.5827	1.5223
PURCHASES_TRX	71.1865	7.7203	9.9169

Cluster 0 dapat diidentifikasi sebagai nasabah aktif dalam transaksi pembelian. Hal ini terlihat dari tingginya nilai rata-rata pada atribut INSTALLMENTS_PURCHASES, PURCHASES_FREQUENCY, dan PURCHASES_TRX. Selain itu, frekuensi pembelian satu kali dan cicilan juga tinggi, menunjukkan bahwa nasabah dalam kelompok ini memanfaatkan kartu kredit secara aktif untuk berbagai jenis pembelian. Sebaliknya, nilai CASH_ADVANCE dan frekuensinya relatif rendah, mengindikasikan bahwa mereka jarang menggunakan kartu kredit untuk tarik tunai. Karakteristik ini menunjukkan perilaku nasabah yang cenderung menggunakan kartu kredit sebagai alat pembayaran utama, bukan sebagai sumber dana tunai.

Cluster 1 menunjukkan pola yang sangat berbeda dan dapat dikategorikan sebagai nasabah dengan preferensi tinggi terhadap tarik tunai. Rata-rata nilai CASH_ADVANCE, CASH_ADVANCE_FREQUENCY, dan CASH_ADVANCE_TRX sangat tinggi dibandingkan dua cluster lainnya. Sementara itu, aktivitas pembelian sangat rendah, baik dari sisi frekuensi maupun jumlah transaksi. Pola ini menunjukkan bahwa nasabah dalam *cluster* ini lebih memanfaatkan kartu kredit untuk kebutuhan likuiditas jangka pendek daripada untuk konsumsi barang atau jasa.

Cluster 2 merepresentasikan nasabah dengan aktivitas transaksi moderat. Nilai-nilai rata-rata pada semua atribut berada di antara cluster 0 dan 1. Mereka memiliki frekuensi pembelian dan penggunaan tarik tunai yang sedang, tanpa kecenderungan dominan ke salah satu jenis transaksi. Hal ini menunjukkan bahwa

nasabah dalam cluster ini menggunakan kartu kredit secara seimbang, baik untuk pembelian maupun tarik tunai, meskipun tidak seintensif cluster 0 dalam pembelian atau cluster 1 dalam tarik tunai.

Visualisasi hasil clustering menggunakan PCA dua dimensi pada Gambar 7 juga memperkuat interpretasi ini, di mana ketiga cluster tampak terpisah dengan cukup jelas. Hal ini menunjukkan bahwa kombinasi fitur yang digunakan dalam proses clustering berhasil membedakan pola transaksi nasabah secara efektif. Dengan demikian, hasil segmentasi ini dapat menjadi dasar yang kuat untuk pengembangan strategi pemasaran yang lebih personal dan tepat sasaran, seperti penawaran produk cicilan untuk cluster 0, program pengelolaan keuangan untuk cluster 1, dan promosi fleksibel untuk cluster 2.

5. Simpulan

Penelitian ini berhasil melakukan segmentasi nasabah kartu kredit berdasarkan pola transaksi dengan pendekatan yang diperbarui, khususnya pada tahap standarisasi data. Penggunaan metode RobustScaler sebagai pengganti StandardScaler terbukti memberikan hasil yang lebih stabil dalam menghadapi outlier yang umum ditemukan pada data transaksi keuangan. Dengan pendekatan ini, segmentasi yang dihasilkan menjadi lebih representatif dan akurat.

Faktor-faktor yang memengaruhi keberhasilan segmentasi dalam penelitian ini antara lain adalah pemilihan fitur yang relevan berdasarkan domain knowledge, metode standarisasi yang sesuai dengan karakteristik data, serta evaluasi model yang sistematis. Kombinasi fitur yang digunakan mencerminkan dimensi penting dari pola transaksi nasabah, seperti frekuensi pembelian, penggunaan fasilitas tarik tunai, dan kemampuan membayar.

Secara praktis, hasil segmentasi ini dapat dimanfaatkan oleh institusi keuangan untuk menyusun strategi pemasaran yang lebih personal dan efisien. Misalnya, nasabah dalam cluster aktif pembelian dapat ditargetkan dengan program cicilan atau reward belanja, sementara nasabah dengan preferensi tarik tunai dapat diberikan edukasi finansial atau penawaran produk pinjaman yang lebih sesuai.

Untuk penelitian selanjutnya, disarankan agar model segmentasi diperluas dengan mempertimbangkan atribut tambahan seperti data demografis, riwayat kredit, atau perilaku digital nasabah. Selain itu, validasi eksternal melalui implementasi strategi berbasis cluster di dunia nyata juga penting dilakukan untuk mengukur efektivitas segmentasi secara praktis.

Pustaka

[1] F. P. Rachman, H. Santoso, and A. Djajadi, "Machine learning mini batch k-means and business intelligence utilization for credit card customer segmentation," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 10, pp. 218–227, 2021.

[2] A. R. Savira, A. M. Siregar, and D. S. Kusumaningrum, "Optimizing clustering models using principle component analysis for car customers," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 18, no. 2, pp. 1–12, Jul 2024.

[3] A. P. Wicaksono, S. Widjaja, M. F. Nugroho, and P. Putri, "Analisa nilai k terhadap pengelompokkan penjualan tiket pada k-means dengan metode elbow dan silhouette," Jan 2024.

[4] T. Tavor, L. D. Gonen, and U. Spiegel, "Customer segmentation as a revenue generator for profit purposes," *Mathematics*, vol. 11, no. 21, Nov 2023.

[5] M. M. Rahman, "The effect of business intelligence on bank operational efficiency and perceptions of profitability," *FinTech*, vol. 2, no. 1, pp. 99–119, Feb 2023.

[6] C. S. Potluri, G. S. Rao, L. V. R. M. Kumar, K. G. Allo, Y. Awoke, and A. A. Seman, "Machine learning-based customer segmentation and personalised marketing in financial services," in *Proceedings of International Conference on Communication, Computer Sciences and Engineering (IC3SE 2024)*. IEEE, Jul 2024, pp. 1570–1574.

[7] S. Nofal, "Identifying highly-valued bank customers with current accounts based on the frequency and amount of transactions," *Heliyon*, vol. 10, no. 13, 2024.

[8] A. Ullah *et al.*, "Customer analysis using machine learning-based classification algorithms for effective segmentation using recency, frequency, monetary, and time," *Sensors*, vol. 23, no. 6, p. 3180, 2023.

[9] S. Monalisa, Y. Juniarti, E. Saputra, F. Muttakin, and T. K. Ahsyar, "Customer segmentation with rfm models and demographic variable using dbscan algorithm," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 21, no. 4, pp. 742–749, 2023.

[10] A. P. Bagustio, A. I. Purnamasari, and I. Ali, "Analisis data penjualan menggunakan algoritma k-means clustering pada toko kecantikan putri," *PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer*, vol. 11, no. 2, pp. 159–167, 2024.

[11] S. Dwididanti and D. A. Anggoro, "Analisis perbandingan algoritma bisecting k-means dan fuzzy c-means pada data pengguna kartu kredit," *Emitor: Jurnal Teknik Elektro*, vol. 22, no. 2, pp. 110–117, 2022.

[12] H. Ramadhan, M. R. A. Kamaludin, M. A. Nasrullah, and D. Rolliawati, "Comparison of hierarchical, k-means and dbscan clustering methods for credit card customer segmentation analysis based on expenditure level," *Journal of Applied Informatics and Computing*, vol. 7, no. 2, pp. 246–251, 2023.

[13] I. M. K. Karo, "Segmentation of credit card customers based on their credit card usage behavior using the k-means algorithm," *Journal of Software Engineering, Information and Communication Technology (SEICT)*, vol. 2, no. 2, pp. 55–64, Feb 2022.

[14] F. D. S. Alhamdani, A. A. Dianti, and Y. Azhar, "Segmentasi pelanggan berdasarkan perilaku penggunaan kartu kredit menggunakan metode k-means clustering," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 6, no. 2, pp. 70–77, May 2021.

[15] S. D. K. Wardani, A. S. Ariyanto, M. Umroh, and D. Rolliawati, "Perbandingan hasil metode clustering k-means, db scanner & hierarchical untuk analisa segmentasi pasar," *JIKO (Jurnal Informatika dan Komputer)*, vol. 7, no. 2, p. 191, Sep 2023.

[16] N. H. M. M. Shrifan, M. F. Akbar, and N. A. M. Isa, "An adaptive outlier removal aided k-means clustering algorithm," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 8, pp. 6365–6376, Sep 2022.

[17] M. M. Ahsan, M. A. P. Mahmud, P. K. Saha, K. D. Gupta, and Z. Siddique, "Effect of data scaling methods on machine learning algorithms and model performance," *Technologies (Basel)*, vol. 9, no. 3, Sep 2021.

[18] G. G. Hamedani *et al.*, "A new right-skewed one-parameter distribution with mathematical characterizations, distributional validation, and actuarial risk analysis, with applications," *Symmetry (Basel)*, vol. 15, no. 7, Jul 2023.

[19] N. P. Santos, "The expansion of data science: Dataset standardization," *Standards*, vol. 3, no. 4, pp. 400–410, Nov 2023.

[20] M. Goyal and Q. H. Mahmoud, "A systematic review of synthetic data generation techniques using generative ai," Sep 2024.

[21] P. Martins, F. Cardoso, P. Váz, J. Silva, and M. Abbasi, "Performance and scalability of data cleaning and preprocessing tools: A benchmark on large real-world datasets," *Data (Basel)*, vol. 10, no. 5, May 2025.

[22] M. Shantal, Z. Othman, and A. A. Bakar, "A novel approach for data feature weighting using correlation coefficients and min–max normalization," *Symmetry (Basel)*, vol. 15, no. 12, Dec 2023.

[23] Z. Hou, J. Liu, Z. Shao, Q. Ma, and W. Liu, "Machine learning innovations in renewable energy systems with integrated nrbo-txad for enhanced wind speed forecasting accuracy," *Electronics (Switzerland)*, vol. 14, no. 12, Jun 2025.