

Segmentasi Pelanggan Kartu Kredit Menggunakan Metode Klustering: Analisis dan Profiling

Agus Arifudin*, Fikri Budiman

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

Email: ¹111202113736@mhs.dinus.ac.id, ²Fikri.budiman@dsn.dinus.ac.id

Email Penulis Korespondensi: 111202113736@mhs.dinus.ac.id

Submitted: 31/01/2025; Accepted: 05/03/2025; Published: 07/03/2025

Abstrak—Penggunaan kartu kredit di Indonesia meningkat signifikan, menciptakan tantangan kompleks bagi lembaga keuangan dalam memahami perilaku pengguna dan memenuhi kebutuhan mereka. Peningkatan ini berpotensi menyebabkan risiko penipuan lebih tinggi, ketidakpuasan pelanggan akibat harapan tidak terpenuhi, serta ketidakstabilan keuangan bagi konsumen dan bank. Hal ini menekankan urgensi penelitian untuk melakukan segmentasi pelanggan berdasarkan perilaku penggunaan mereka. Dataset yang dianalisis mencakup informasi dari 8.950 pengguna kartu kredit, termasuk frekuensi transaksi, jumlah saldo, dan jenis transaksi. Penelitian ini bertujuan melakukan segmentasi pelanggan menggunakan algoritma *K-Means*, *DBSCAN*, dan *Hierarchical Clustering*. *K-Means* mengelompokkan pelanggan dengan pola perilaku serupa, *DBSCAN* mengidentifikasi kluster tidak beraturan dan *outlier*, sementara *Hierarchical Clustering* memberikan wawasan tentang hubungan antar kluster. Hasil analisis menunjukkan empat segmen utama dengan karakteristik unik. Misalnya, segmen pengguna aktif memiliki frekuensi transaksi tinggi dan saldo besar, sedangkan segmen pengguna baru menunjukkan frekuensi transaksi lebih rendah. Temuan ini memberikan wawasan berharga bagi lembaga keuangan untuk meningkatkan layanan dan penawaran produk. Dengan memahami karakteristik setiap segmen, lembaga keuangan dapat menyesuaikan strategi pemasaran dan produk untuk meningkatkan kepuasan pelanggan dan loyalitas.

Kata Kunci: Segmentasi Pelanggan; Kartu Kredit; K-Means; DBSCAN; Hierarchical Clustering

Abstract—The use of credit cards in Indonesia has increased significantly, creating complex challenges for financial institutions in understanding user behavior and meeting their needs. This growth poses a higher risk of fraud, customer dissatisfaction due to unmet expectations, and financial instability for both consumers and banks. These issues highlight the urgency of conducting research to segment customers based on their usage behavior. The analyzed dataset includes information from 8,950 credit card users, covering transaction frequency, account balance, and transaction types. This study aims to segment customers using K-Means, DBSCAN, and Hierarchical Clustering algorithms. K-Means groups customers with similar behavioral patterns, DBSCAN identifies irregular clusters and outliers, while Hierarchical Clustering provides insights into relationships between clusters. The analysis results reveal four main segments, each with unique characteristics. For instance, the active user segment exhibits high transaction frequency and large balances, whereas new users demonstrate lower transaction frequency. These findings offer valuable insights for financial institutions to enhance their services and product offerings. By understanding the characteristics of each segment, financial institutions can tailor their marketing strategies and products to improve customer satisfaction and loyalty.

Keywords: Segmentation; Credit Card; K-Means; DBSCAN; Hierarchical Clustering

1. PENDAHULUAN

Kartu kredit telah menjadi salah satu instrumen pembayaran yang paling umum digunakan di seluruh dunia, termasuk di Indonesia. Dengan kemudahan yang ditawarkannya, kartu kredit memungkinkan konsumen untuk melakukan transaksi secara cepat dan efisien, serta memberikan akses ke berbagai layanan keuangan yang tidak tersedia melalui metode pembayaran tradisional. Menurut data dari Bank Indonesia, penggunaan kartu kredit di Indonesia terus meningkat, dengan jumlah pengguna yang mencapai jutaan orang dan transaksi yang signifikan setiap tahunnya. Hal ini menunjukkan bahwa kartu kredit tidak hanya berfungsi sebagai alat pembayaran, tetapi juga sebagai instrumen keuangan yang penting dalam kehidupan sehari-hari masyarakat[1].

Namun, dengan meningkatnya jumlah pengguna kartu kredit, tantangan dalam memahami perilaku dan preferensi pelanggan juga semakin kompleks. Penelitian oleh Robb[2] menunjukkan bahwa pengetahuan keuangan memiliki pengaruh signifikan terhadap perilaku penggunaan kartu kredit di kalangan mahasiswa. Penelitian ini menekankan bahwa individu dengan pengetahuan keuangan yang lebih baik cenderung menggunakan kartu kredit secara lebih bertanggung jawab, yang menunjukkan pentingnya pemahaman tentang faktor-faktor yang mempengaruhi perilaku konsumen. Meskipun pihak perbankan telah menyediakan fitur keamanan yang canggih, masih terdapat celah yang dapat dimanfaatkan oleh penipu untuk melakukan kejahatan. Oleh karena itu, penting bagi lembaga keuangan untuk memahami perilaku pelanggan dan mengembangkan strategi yang dapat meningkatkan keamanan serta kepuasan pelanggan.

Segmentasi pelanggan merupakan pendekatan yang efektif untuk mengidentifikasi kelompok-kelompok pelanggan dengan karakteristik dan perilaku yang serupa. Dengan memahami segmen-segmen ini, lembaga keuangan dapat merancang produk dan layanan yang lebih sesuai dengan kebutuhan spesifik masing-masing kelompok, yang pada gilirannya dapat meningkatkan kepuasan pelanggan dan loyalitas. Penelitian sebelumnya menunjukkan bahwa segmentasi yang tepat dapat meningkatkan efektivitas kampanye pemasaran dan memaksimalkan keuntungan[3]. Selain itu, penelitian oleh Wicaksono[4] menekankan pentingnya segmentasi pelanggan dalam mengembangkan strategi pemasaran yang disesuaikan dengan karakteristik pelanggan.

Dalam konteks ini, analisis klustering menjadi alat yang sangat berguna. Klustering adalah teknik yang digunakan untuk mengelompokkan data ke dalam grup yang lebih kecil berdasarkan kesamaan di antara mereka[5]. Metode klustering memungkinkan perusahaan untuk mengidentifikasi pola dalam perilaku pelanggan yang mungkin tidak terlihat melalui analisis tradisional. Dengan menggunakan metode klustering, perusahaan dapat mengidentifikasi segmen pelanggan yang berbeda dan merancang penawaran yang lebih sesuai dengan kebutuhan masing-masing segmen. Beberapa metode klustering yang umum digunakan termasuk *K-Means*, *DBSCAN*, dan *Hierarchical Clustering*, masing-masing dengan kelebihan dan kekurangan tersendiri[6][7].

K-Means adalah salah satu metode klustering yang paling banyak digunakan karena kesederhanaannya dan kemampuannya dalam menangani dataset besar. Metode ini bekerja dengan mengelompokkan data ke dalam k kluster berdasarkan jarak ke *centroid* (titik pusat) dari setiap kluster. Proses ini melibatkan pemilihan jumlah kluster yang diinginkan dan iterasi untuk meminimalkan jarak antara data dan centroid[3][5][8][9][10]. Sementara itu, *DBSCAN* (*Density-Based Spatial Clustering of Applications with Noise*) adalah metode yang lebih canggih yang dapat mengidentifikasi kluster dengan bentuk yang tidak teratur dan juga dapat mendeteksi *outlier*[6][11][12]. Metode ini sangat berguna dalam situasi di mana data tidak terdistribusi secara merata. Di sisi lain, *Hierarchical Clustering* menawarkan pendekatan yang berbeda dengan membangun struktur pohon (*dendrogram*) untuk mengelompokkan data, memberikan wawasan yang lebih mendalam tentang hubungan antar kluster [5][7].

Meskipun penelitian-penelitian sebelumnya memberikan kontribusi yang signifikan, masih terdapat celah dalam pemahaman perilaku pengguna kartu kredit secara spesifik di Indonesia, terutama dalam konteks penggunaan algoritma klustering yang berbeda. Penelitian oleh Wicaksono menunjukkan bahwa penggunaan algoritma klustering dapat membantu dalam segmentasi pelanggan berdasarkan perilaku dan preferensi mereka[4]. Selain itu, penelitian oleh Karo Karo et al[1]. menekankan pentingnya analisis klustering dalam memahami karakteristik pelanggan yang berbeda, yang dapat meningkatkan efektivitas strategi pemasaran. Penelitian ini menunjukkan bahwa penerapan metode klustering yang tepat, seperti *K-Means*, dapat membantu dalam mengidentifikasi segmen pasar yang lebih spesifik dan relevan, serta memberikan wawasan yang lebih baik bagi lembaga keuangan dalam merancang produk dan layanan yang sesuai dengan kebutuhan pelanggan[6][10][11][13].

Lebih lanjut, penelitian oleh Rohalidyawati et al. [13] menunjukkan bahwa algoritma *DBSCAN* dapat digunakan untuk segmentasi pelanggan *e-money*, yang memberikan wawasan tambahan tentang perilaku pengguna dalam konteks yang lebih luas. Penelitian ini menunjukkan bahwa pendekatan berbasis kepadatan dapat mengungkap pola yang tidak terdeteksi oleh metode klustering lainnya. Selain itu, Abednego et al. [11] menyoroti pentingnya strategi pemasaran yang ditingkatkan melalui penggunaan algoritma *K-Means*, yang dapat membantu dalam merancang kampanye yang lebih efektif. Penelitian oleh Pranata et al. [14] juga menunjukkan bahwa penerapan algoritma *K-Means* dalam sektor perbankan dapat meningkatkan pemahaman tentang perilaku pelanggan dan meningkatkan retensi. Di samping itu, Sarkar et al. [15] membahas pentingnya segmentasi dalam konteks kartu kredit, yang relevan dengan penelitian ini. Penelitian oleh Zhao et al. [16] menekankan penggunaan *mini-batch K-Means* untuk meningkatkan efisiensi dalam segmentasi pelanggan, sedangkan Fadhillah et al. [5] menunjukkan bagaimana atribut *Recency*, *Frequency*, dan *Monetary (RFM)* dapat digunakan dalam analisis klustering untuk meningkatkan pemahaman tentang pelanggan. Terakhir, Paramita dan Hariguna[6] membandingkan efektivitas algoritma *K-Means* dan *DBSCAN* dalam segmentasi pelanggan, memberikan wawasan yang berharga untuk penelitian ini.

Dari riset-riset sebelumnya, meskipun banyak yang memberikan kontribusi signifikan terhadap pemahaman perilaku pengguna kartu kredit, belum dilakukan analisis yang komprehensif mengenai segmentasi pelanggan kartu kredit di Indonesia dengan menggunakan berbagai algoritma klustering secara bersamaan. Pada penelitian ini, kami melakukan eksperimen dengan menerapkan algoritma *K-Means*, *DBSCAN*, dan *Hierarchical Clustering* untuk menganalisis dataset yang terdiri dari informasi transaksi dari sekitar 8.950 pengguna kartu kredit. Dataset ini mencakup frekuensi transaksi, jumlah pembelian, dan saldo yang tersisa, yang memungkinkan kami untuk mendapatkan wawasan yang lebih mendalam tentang perilaku pengguna. Dengan demikian, penelitian ini bertujuan untuk mengisi celah yang ada dalam literatur dengan menerapkan metode klustering yang berbeda untuk segmentasi pelanggan kartu kredit di Indonesia. Penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam tentang karakteristik segmen pelanggan dan membantu lembaga keuangan dalam merancang produk dan layanan yang lebih sesuai dengan kebutuhan mereka. Dengan memahami perilaku dan preferensi pelanggan, lembaga keuangan dapat meningkatkan strategi pemasaran mereka dan pada akhirnya meningkatkan kepuasan dan loyalitas pelanggan[17][18]. Untuk menguji hasil klustering yang diperoleh, kami menggunakan metrik evaluasi seperti *indeks Davies-Bouldin*, *skor silhouette*, dan *indeks Calinski-Harabasz*, yang akan memberikan gambaran tentang kualitas kluster yang terbentuk.

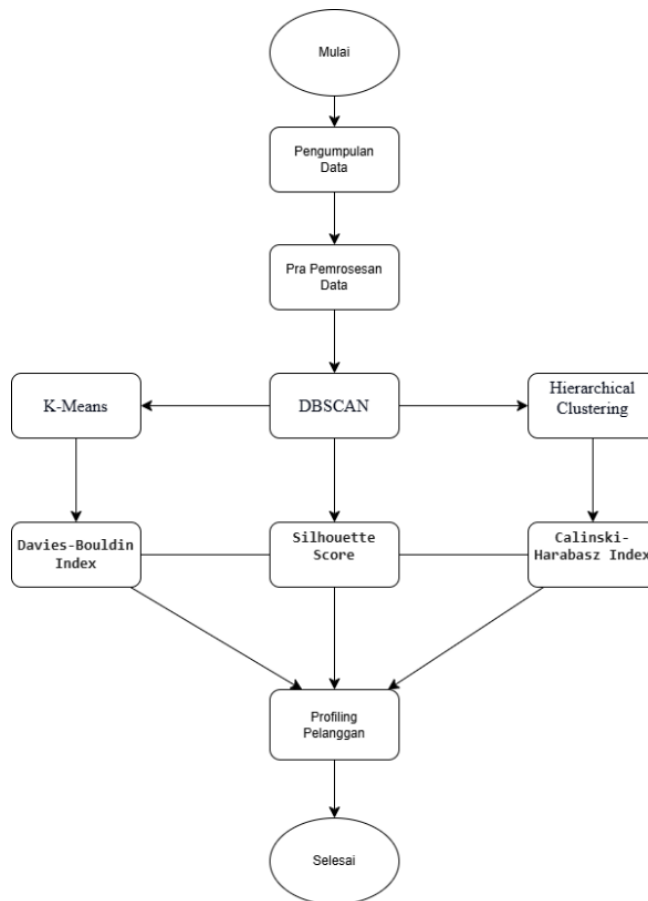
Dalam konteks ini, penelitian ini diharapkan dapat memberikan kontribusi signifikan terhadap pemahaman perilaku pengguna kartu kredit di Indonesia. Dengan menerapkan berbagai algoritma klustering, penelitian ini tidak hanya bertujuan untuk mengidentifikasi segmen-segmen pelanggan yang berbeda, tetapi juga untuk memberikan wawasan yang lebih mendalam tentang karakteristik dan preferensi masing-masing segmen. Hasil dari penelitian ini diharapkan dapat menjadi referensi bagi lembaga keuangan dalam merancang strategi pemasaran yang lebih efektif dan produk yang lebih sesuai dengan kebutuhan pelanggan. Selain itu, penelitian ini juga berpotensi untuk memperkaya literatur di bidang analisis klustering, khususnya dalam konteks aplikasi di sektor perbankan, dengan memberikan contoh konkret tentang bagaimana teknik-teknik ini dapat digunakan untuk meningkatkan pengalaman pelanggan dan efisiensi operasional. Dengan demikian, penelitian ini tidak hanya berfokus pada aspek bisnis, tetapi

juga pada peningkatan pengalaman pelanggan secara keseluruhan, yang pada akhirnya dapat berkontribusi pada pertumbuhan dan stabilitas industri perbankan di Indonesia.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada bagian ini, akan dijelaskan tahapan-tahapan yang dilakukan dalam penelitian ini untuk mencapai tujuan segmentasi pelanggan kartu kredit. Setiap tahapan dirancang secara sistematis untuk memastikan bahwa proses penelitian berjalan dengan efektif dan efisien. Gambar 1 di bawah ini menggambarkan alur penelitian secara keseluruhan, mulai dari pengumpulan data hingga profiling pelanggan, serta langkah-langkah yang diambil hingga proses penelitian dinyatakan selesai.



Gambar 1. Tahapan penelitian

Berdasarkan alur penelitian pada Gambar 1, Penelitian ini bertujuan untuk melakukan segmentasi pelanggan kartu kredit melalui serangkaian tahapan yang sistematis. Proses dimulai dengan pengumpulan data pelanggan dari sumber yang relevan, mencakup informasi seperti batas kredit, saldo, dan frekuensi transaksi. Setelah data terkumpul, tahap pra-pemrosesan dilakukan untuk memastikan kualitas data, termasuk pembersihan dan normalisasi. Selanjutnya, metode klustering digunakan dipilih, yaitu *K-Means*, *DBSCAN*, dan *Hierarchical Clustering*. Setelah metode ditentukan, klustering diterapkan pada data yang telah diproses, menghasilkan kluster yang berbeda. Untuk mengevaluasi kualitas kluster, beberapa indeks digunakan, seperti *Davies-Bouldin Index*, *Silhouette Score*, dan *Calinski-Harabasz Index*[6][7][12][14][19].

Hasil klustering kemudian dianalisis untuk membuat profil pelanggan, yang mencakup karakteristik dan perilaku dari setiap segmen yang diidentifikasi. Proses penelitian diakhiri dengan penyusunan laporan yang mencakup temuan dan rekomendasi berdasarkan hasil analisis. Dengan pendekatan ini, penelitian diharapkan dapat memberikan wawasan yang lebih mendalam tentang karakteristik segmen pelanggan kartu kredit dan membantu lembaga keuangan dalam merancang produk yang lebih sesuai dengan kebutuhan pelanggan[10][20][21][22].

2.2 Pengumpulan Data

Data yang digunakan adalah dataset publik tentang *credit card dataset* yang diperoleh dari *Kaggle*, yang dapat diakses melalui www.kaggle.com. Dataset yang di ambil berjumlah 8.950 data dengan 18 atribut dan 2 kelas. Tabel 1 merupakan spesifikasi atribut beserta jenis datanya.

Tabel 1. Atribut dan Tipe Data

Atribut	Tipe Data	Keterangan
ID_Pelanggan	Teks	Nomor ID Pelanggan
Saldo	Numerik	Saldo akun yang tersedia untuk pembelian
Frekuensi_Saldo	Numerik	Frekuensi pembaruan saldo
Pembelian	Numerik	Jumlah pembelian akun
Pembelian_Sekali	Numerik	Jumlah maksimum pembelian dalam satu transaksi
Pembelian_Cicilan	Numerik	Jumlah pembelian dalam cicilan
Uang_Muka	Numerik	Pembayaran uang muka
Frekuensi_Pembelian	Numerik	Frekuensi pembelian yang dilakukan secara reguler
Frekuensi_Pembelian_Sekali	Numerik	Frekuensi pembelian yang dilakukan dalam satu transaksi
Frekuensi_Pembelian_Cicilan	Numerik	Frekuensi pembelian yang dilakukan dalam cicilan
Frekuensi_Uang_Muka	Numerik	Frekuensi uang muka
Transaksi_Uang_Muka	Numerik	Total transaksi uang muka
Transaksi_Pembelian	Numerik	Total transaksi pembelian
Batas_Kredit	Numerik	Batas kredit pengguna
Pembayaran	Numerik	Total jumlah yang dibayarkan oleh pengguna
Pembayaran_Minimum	Numerik	Jumlah pembayaran minimum yang dilakukan oleh pengguna
Pembayaran_Penuh	Numerik	Persentase total tagihan yang dibayarkan oleh pengguna
Tenur	Numerik	Tenur kartu kredit pengguna

2.3 Pra-pemrosesan Data

Tahap pra-pemrosesan data merupakan langkah krusial dalam analisis klustering, bertujuan untuk memastikan kualitas dan konsistensi data sebelum diterapkan metode klustering. Berikut adalah langkah-langkah yang dilakukan dalam pra-pemrosesan data untuk penelitian ini:

- Menghapus Variabel yang Tidak Diperlukan: Variabel yang tidak relevan atau tidak memberikan informasi yang berguna untuk analisis klustering dihapus dari dataset untuk mengurangi kompleksitas dan meningkatkan efisiensi.
- Imputasi: Nilai yang hilang dalam dataset diisi menggunakan teknik *imputasi*, seperti imputasi rata-rata untuk variabel numerik atau modus untuk variabel kategorikal, untuk memastikan bahwa analisis tidak terpengaruh oleh data yang hilang.
- Normalisasi Data: Data dinormalisasi menggunakan metode *Min-Max Scaling* atau *Z-score* untuk memastikan bahwa semua variabel berada dalam skala yang sama, sehingga tidak ada variabel yang mendominasi hasil klustering.
- Uji Hopkins: Uji Hopkins dilakukan untuk mengukur sejauh mana data memiliki kecenderungan untuk membentuk kluster. Nilai Hopkins yang mendekati 1 menunjukkan bahwa data memiliki struktur kluster yang kuat.
- Analisis Komponen Utama (PCA): *PCA* diterapkan untuk mereduksi dimensi data, sehingga memudahkan visualisasi dan analisis. Dengan mengurangi dimensi, *PCA* membantu mengidentifikasi pola yang mungkin tidak terlihat dalam data berdimensi tinggi.

2.4 Penerapan Klustering

Setelah pra-pemrosesan, langkah selanjutnya adalah penerapan teknik klustering. Tiga metode klustering yang digunakan dalam penelitian ini adalah:

- K-Means Clustering*: *K-Means* digunakan untuk mengelompokkan data berdasarkan kesamaan. Proses ini melibatkan inisialisasi *centroid* secara acak, penugasan titik data ke *centroid* terdekat, dan pembaruan *centroid* hingga *konvergensi*. *K-Means* dipilih karena kemampuannya untuk mengelompokkan data ke dalam kluster berdasarkan kesamaan perilaku. Metode ini efektif untuk dataset besar dan dapat memberikan hasil yang cepat. Namun, *K-Means* memiliki kelemahan dalam menentukan jumlah kluster yang optimal dan dapat sensitif terhadap outlier, yang dapat mempengaruhi hasil klustering. Rumus penugasan titik data x_i ke cluster j adalah:

$$C_j = \{x_i: ||x_i - \mu_j||^2 \leq ||x_i - \mu_k||^2, \forall k \neq j\} \quad (1)$$

Setelah itu, *centroid* diperbarui dengan rumus:

$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \quad (2)$$

Jumlah kluster yang optimal ditentukan menggunakan metode *elbow* dan *Calinski-Harabasz index*.

- DBSCAN (Density-Based Spatial Clustering of Applications with Noise)*: *DBSCAN* digunakan untuk mengidentifikasi kluster dengan bentuk yang tidak teratur dan mendeteksi *outlier*. Parameter *MinPoints* dan *Epsilon* ditentukan sebelum menerapkan algoritma. Titik diklasifikasikan sebagai *core point*, *border point*, atau *noise point* berdasarkan kepadatan. Metode ini sangat berguna dalam konteks pelanggan kartu kredit, di mana pola transaksi dapat bervariasi secara signifikan. Kelebihan *DBSCAN* adalah kemampuannya untuk menemukan

kluster dengan kepadatan yang berbeda dan tidak memerlukan penentuan jumlah kluster sebelumnya. Namun, DBSCAN dapat kesulitan dalam mengidentifikasi kluster yang memiliki kepadatan yang sangat berbeda.

- c. *Hierarchical Clustering (Agglomerative)*: Metode ini memberikan wawasan tambahan tentang hubungan antar kluster dengan membangun struktur pohon (dendrogram) yang menunjukkan bagaimana kluster terbentuk. Metode ini memungkinkan analisis yang lebih mendalam tentang hierarki dalam data pelanggan. Meskipun memberikan informasi yang lebih kaya, *Hierarchical Clustering* dapat menjadi kurang efisien untuk dataset yang sangat besar dan dapat memerlukan waktu komputasi yang lebih lama.. Jarak antar titik dihitung menggunakan rumus jarak *Euclidean*:

$$d(A, B) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \tag{3}$$

2.5 Evaluasi

- a. *Davies-Bouldin Index*: Mengukur rasio antara jarak intra-cluster dan jarak antar-cluster. Nilai yang lebih rendah menunjukkan kluster yang lebih baik.
- b. *Silhouette Score*: Mengukur seberapa dekat setiap titik data dengan kluster yang benar dibandingkan dengan kluster terdekat lainnya. Nilai berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan kluster yang lebih baik.
- c. *Calinski-Harabasz Index*: Mengukur rasio antara varians antar-kluster dan varians intra-cluster. Nilai yang lebih tinggi menunjukkan kluster yang lebih baik.

2.6 Analisis dan Profiling

Analisis dalam konteks penelitian ini merujuk pada proses evaluasi data yang telah dikumpulkan untuk mengidentifikasi pola, tren, dan karakteristik yang relevan. Dalam penelitian segmentasi pelanggan kartu kredit, analisis dilakukan setelah penerapan metode klustering untuk memahami bagaimana pelanggan berperilaku berdasarkan atribut yang telah ditentukan. Proses ini melibatkan penggunaan berbagai teknik statistik dan visualisasi untuk mengevaluasi hasil klustering dan menilai kualitas segmen yang terbentuk. *Profiling* adalah langkah selanjutnya setelah analisis, di mana karakteristik setiap segmen pelanggan diidentifikasi dan dijelaskan secara mendetail. *Profiling* mencakup pembuatan deskripsi yang komprehensif tentang setiap kluster, termasuk informasi demografis, perilaku keuangan, dan preferensi penggunaan kartu kredit. Dengan profiling, lembaga keuangan dapat memahami kebutuhan spesifik dari setiap segmen, yang memungkinkan mereka untuk merancang produk dan layanan yang lebih sesuai.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Preprocessing

Tahap pra-pemrosesan data merupakan langkah krusial dalam analisis klustering, bertujuan untuk memastikan kualitas dan konsistensi data sebelum diterapkan metode klustering. Berikut adalah langkah-langkah yang dilakukan dalam pra-pemrosesan data untuk penelitian ini:

- a. Menghapus Variabel yang Tidak Diperlukan

Variabel *ID_Pelanggan* dihapus dari dataset karena memiliki nilai unik untuk setiap pelanggan. Nilai unik ini tidak memberikan informasi yang relevan untuk proses klustering, sehingga dapat mengganggu analisis. Gambar 2 menunjukkan struktur dataset sesudah penghapusan variabel *ID_Pelanggan*.

	Saldo	Frekuensi_Saldo	Pembelian	Pembelian_Sekali	Pembelian_Cicilan	Uang_Muka
0	40.900749	0.818182	95.40	0.00	95.4	0.000000
1	3202.467416	0.909091	0.00	0.00	0.0	6442.945483
2	2495.148862	1.000000	773.17	773.17	0.0	0.000000
3	1666.670542	0.636364	1499.00	1499.00	0.0	205.788017
4	817.714335	1.000000	16.00	16.00	0.0	0.000000

Gambar 2. Menghapus variabel *ID_Pelanggan*

- b. Imputasi

Nilai yang hilang dalam dataset diimputasi menggunakan metode *KNNImputer*. Metode ini mengisi nilai yang hilang dengan rata-rata dari tetangga terdekat, sehingga menghindari bias dalam hasil klustering. Dengan cara ini, data yang hilang tidak akan mempengaruhi integritas analisis. Hasil dataframe setelah proses imputasi ditunjukkan pada Gambar 3.



```

.: Dataframe after Imputation :.
*****

```

	Saldo	Frekuensi_Saldo	Pembelian	Pembelian_Sekali	Pembelian_Cicilan	Uang_Muka
0	40.900749	0.818182	95.40	0.00	95.4	0.000000
1	3202.467416	0.909091	0.00	0.00	0.0	6442.945483
2	2495.148862	1.000000	773.17	773.17	0.0	0.000000
3	1666.670542	0.636364	1499.00	1499.00	0.0	205.788017
4	817.714335	1.000000	16.00	16.00	0.0	0.000000

Gambar 3. Dataframe setelah di imputasi

c. Normalisasi Data

Normalisasi dilakukan untuk menyelaraskan skala variabel yang berbeda. Proses ini penting agar semua variabel memiliki kontribusi yang setara dalam analisis klustering. Normalisasi dilakukan menggunakan *StandardScaler*, yang menghilangkan rata-rata dan menskalakan data ke varians unit. Hal ini memastikan bahwa variabel dengan skala yang lebih besar tidak mendominasi hasil klustering.

Normalisasi dalam konteks ini dilakukan menggunakan *StandardScaler* dari pustaka *scikit-learn*. *StandardScaler* menghilangkan rata-rata dari setiap fitur dan menskalakan data ke varians unit. Dengan kata lain, setiap fitur akan memiliki rata-rata 0 dan deviasi standar 1 setelah proses normalisasi. Ini membantu dalam mengurangi bias yang mungkin ditimbulkan oleh perbedaan skala antar variabel. Hasil Proses normalisasi data dapat dilihat pada Gambar 4.

	0	1	2	3	4	5	6	\
0	-0.731989	-0.249434	-0.424900	-0.356934	-0.349079	-0.466786	-0.806490	
1	0.786961	0.134325	-0.469552	-0.356934	-0.454576	2.605605	-1.221758	
2	0.447135	0.518084	-0.107668	0.108889	-0.454576	-0.466786	1.269843	
3	0.049099	-1.016953	0.232058	0.546189	-0.454576	-0.368653	-1.014125	
4	-0.358775	0.518084	-0.462063	-0.347294	-0.454576	-0.466786	-1.014125	
	7	8	9	10	11	12	13	\
0	-0.678661	-0.707313	-0.675349	-0.476070	-0.511333	-0.960383	-0.528979	
1	-0.678661	-0.916995	0.573963	0.110074	-0.591796	0.688676	0.818642	
2	2.673451	-0.916995	-0.675349	-0.476070	-0.109020	0.826098	-0.383805	
3	-0.399319	-0.916995	-0.258913	-0.329534	-0.551565	0.826098	-0.598688	
4	-0.399319	-0.916995	-0.675349	-0.476070	-0.551565	-0.905414	-0.364368	
	14	15	16					
0	-0.311729	-0.525551	0.36068					
1	0.087047	0.234227	0.36068					
2	-0.103210	-0.525551	0.36068					
3	0.148440	-0.525551	0.36068					
4	-0.266722	-0.525551	0.36068					

Gambar 4. Normalisasi Data

d. Uji Hopkins

Uji Hopkins diterapkan untuk mengevaluasi apakah data pelanggan kartu kredit menunjukkan pola yang mendukung pembentukan kluster. Hasil uji menunjukkan nilai *Hopkins* sebesar 0.9671, yang mendekati 1. Nilai ini mengindikasikan bahwa data memiliki kecenderungan tinggi untuk membentuk kluster yang bermakna, sehingga tidak terdistribusi secara acak.

Dengan demikian, hasil ini menegaskan bahwa terdapat struktur kluster yang signifikan dalam dataset, yang memberikan dasar yang kuat untuk penerapan metode klustering lebih lanjut, seperti *K-Means*, *DBSCAN*, dan *Hierarchical Clustering*. Kesimpulan dari *Uji Hopkins* ini adalah terima H_0 , yang menyatakan bahwa data memiliki struktur kluster yang bermakna, mendukung tujuan penelitian dalam segmentasi pelanggan kartu kredit.

e. Principal Component Analysis (PCA)

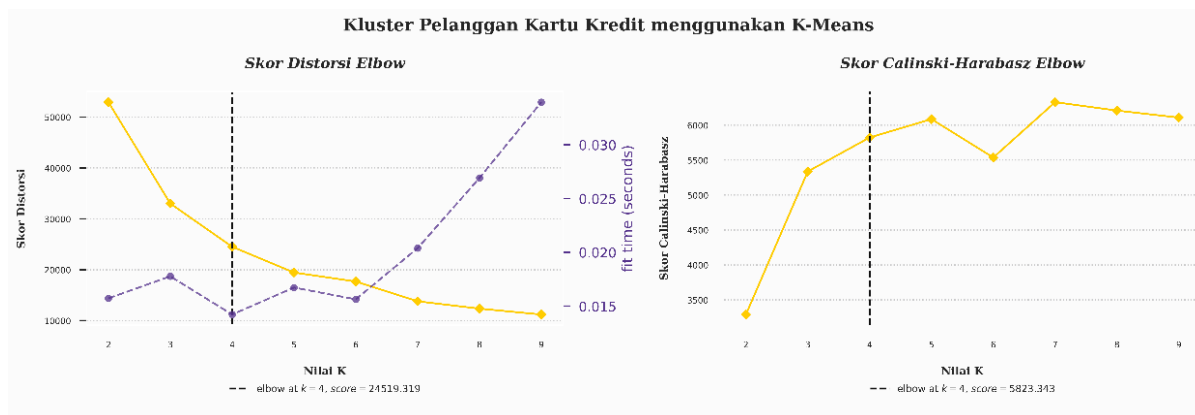
Principal Component Analysis (PCA) atau Analisis Komponen Utama diterapkan dalam penelitian ini untuk mengurangi dimensi data menjadi dua dimensi. Langkah ini bertujuan untuk memudahkan visualisasi hasil klustering dan meningkatkan efisiensi analisis. Dengan mengurangi dimensi, *PCA* membantu dalam mengurangi kompleksitas data, mengurangi *noise*, dan menurunkan biaya komputasi, sehingga memungkinkan analisis yang lebih cepat dan efektif.

Implementasi *PCA* dilakukan dengan menggunakan pustaka *scikit-learn*, di mana data yang telah dinormalisasi diubah menjadi array dan kemudian diterapkan *PCA* dengan parameter $n_components=2$. Proses ini memungkinkan visualisasi yang lebih intuitif dari hasil klustering, sehingga memudahkan dalam mengidentifikasi pola dan struktur dalam data pelanggan kartu kredit.

3.2 Hasil Clustering

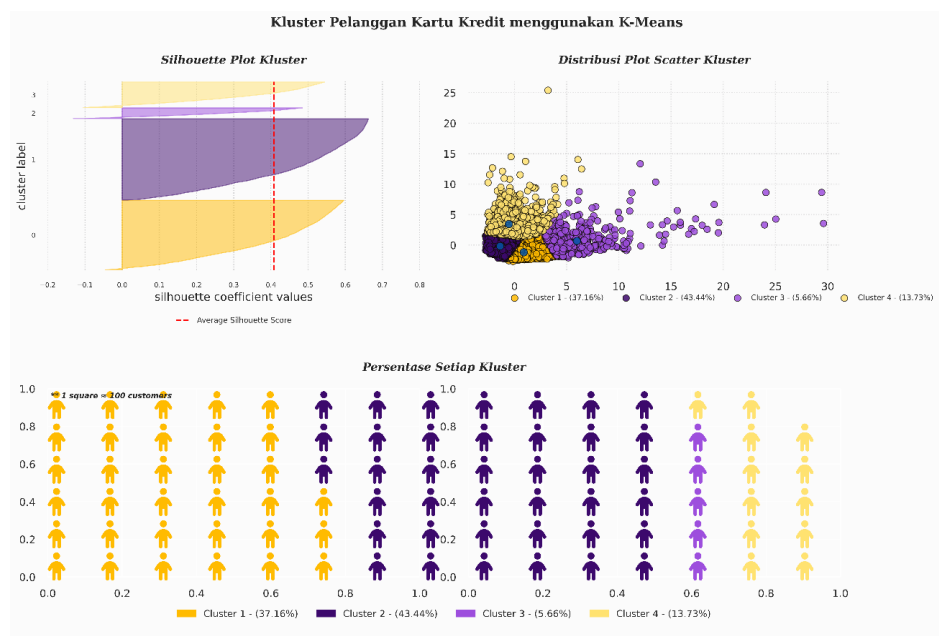
3.2.1 K-Means Clustering

Dalam penelitian ini, algoritma *K-Means* diterapkan untuk melakukan segmentasi pelanggan kartu kredit. Sebelum penerapan kluster, langkah pertama yang dilakukan adalah menghitung jumlah kluster optimal menggunakan metode *elbow* dan *indeks Calinski-Harabasz*. Berdasarkan hasil analisis yang ditunjukkan pada Gambar 5, dapat disimpulkan bahwa jumlah kluster terbaik untuk algoritma *K-Means* adalah 4 kluster.



Gambar 5. Grafik Penentuan Kluster K-Means

Setelah menentukan jumlah kluster optimal, kluster diterapkan dan hasilnya divisualisasikan melalui plot distribusi kluster dan plot *silhouette*, seperti yang ditunjukkan pada Gambar 6. Visualisasi ini menggambarkan nilai koefisien *silhouette* untuk setiap kluster, distribusi setiap kluster dalam plot sebar, serta jumlah pelanggan dalam setiap kluster. Hasil menunjukkan bahwa nilai *silhouette* untuk setiap kluster berada di atas rata-rata, yang menandakan bahwa semua kluster yang terbentuk adalah optimal.



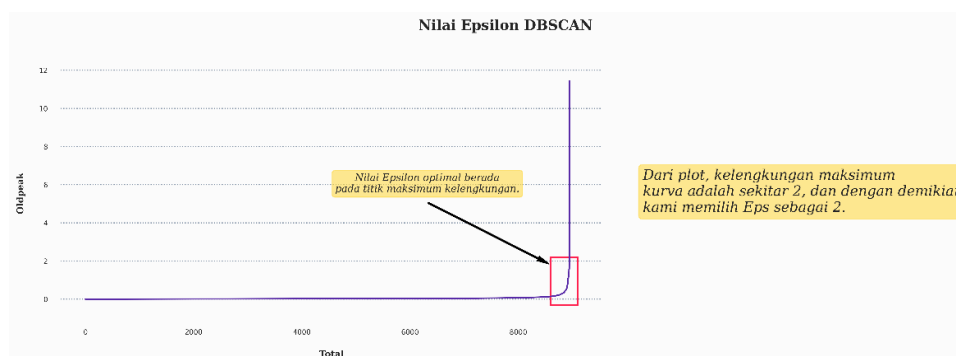
Gambar 6. Hasil Kluster menggunakan K-Means

Dari analisis lebih lanjut, kluster 2 dan 3 menunjukkan viskositas yang lebih tinggi dibandingkan dengan kluster lainnya, dengan lebih dari 35% pelanggan terdistribusi di dalamnya. Hal ini terlihat dari lokasi sebagian besar titik data yang berada di sudut kiri bawah plot sebar. Selain itu, algoritma *K-Means* mengidentifikasi bahwa data outlier termasuk dalam kluster 1 dan 4, di mana outlier pada sumbu x termasuk dalam kluster 1, dan outlier pada sumbu y termasuk dalam kluster 4. *Diagram waffle* yang disertakan dalam visualisasi memberikan gambaran yang jelas tentang persentase distribusi pelanggan dalam setiap kluster, membantu dalam memahami karakteristik dan perilaku segmen pelanggan. Secara keseluruhan, hasil kluster menggunakan *K-Means* menunjukkan bahwa algoritma ini berhasil mengidentifikasi 4 kluster yang optimal, yang dapat digunakan untuk strategi pemasaran dan pengembangan produk yang lebih efektif.

Untuk lebih memahami implikasi dari hasil klustering ini, penting untuk menganalisis kelebihan dan kelemahan dari setiap kluster yang terbentuk. Kelebihan dari hasil klustering K-Means adalah kemampuannya untuk membagi pelanggan menjadi kelompok-kelompok yang relatif homogen berdasarkan karakteristik tertentu, seperti pola penggunaan kartu kredit atau tingkat pengeluaran. Kluster 2 dan 3, yang memiliki viskositas tinggi dengan lebih dari 35% pelanggan, kemungkinan besar mewakili segmen pelanggan yang paling aktif atau loyal. Hal ini dapat menjadi fokus utama dalam strategi pemasaran, seperti penawaran khusus atau program loyalitas. Sementara itu, kluster 1 dan 4, yang mengandung outlier, mungkin mewakili pelanggan dengan perilaku yang tidak biasa atau potensial untuk penipuan, sehingga memerlukan pendekatan khusus seperti pemantauan lebih ketat atau analisis lebih lanjut. Namun, kelemahan dari hasil klustering ini adalah adanya kemungkinan tumpang tindih antara kluster, terutama pada kluster 1 dan 4, yang dapat mengurangi kejelasan segmentasi. Selain itu, K-Means cenderung kurang efektif dalam menangani data yang memiliki bentuk kluster yang tidak teratur atau data dengan noise yang signifikan.

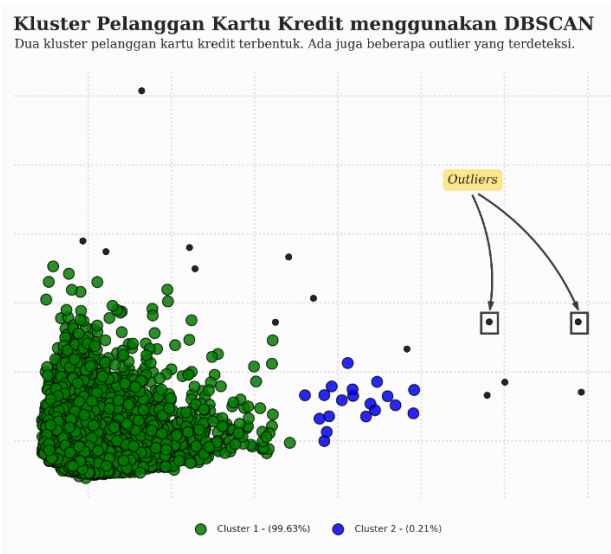
3.2.2 DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Sebelum menerapkan algoritma *DBSCAN*, langkah pertama adalah mendefinisikan parameter-parameter yang diperlukan. Untuk nilai *MinPoints*, karena analisis *PCA* telah dilakukan dan data telah direduksi menjadi dua dimensi, nilai default yang digunakan adalah 4. Untuk nilai *Epsilon*, kami menggunakan algoritma *Nearest Neighbors* untuk menentukan jarak antara setiap titik data dan tetangga terdekatnya, kemudian mengurutkannya sebelum memplotnya. Proses ini membantu dalam menentukan nilai maksimum pada kurva grafik yang dihasilkan, seperti yang ditunjukkan pada Gambar 7.



Gambar 7. Nilai Epsilon DBSCAN

Berdasarkan hasil analisis kurva maksimum dan nilai *MinPoints* yang telah ditentukan sebelumnya, langkah selanjutnya adalah menerapkan *DBSCAN* dan mengevaluasi hasilnya, yang ditampilkan pada Gambar 2.



Gambar 8. Hasil Kluster Menggunakan DBSCAN

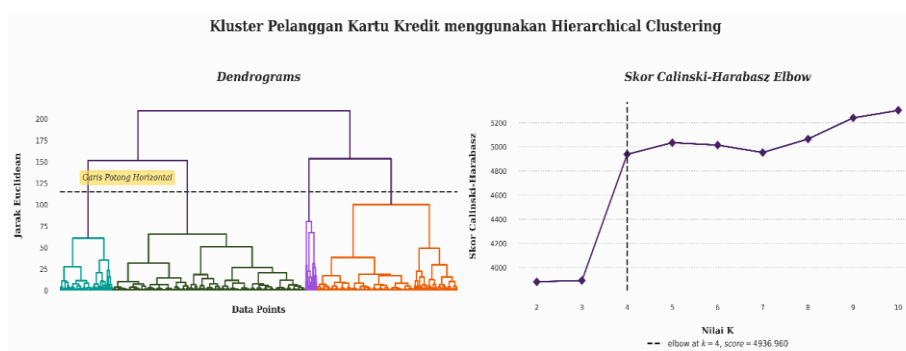
Hasil implementasi *DBSCAN* menunjukkan bahwa terdapat 2 kluster yang terbentuk. Kluster 1 memiliki jumlah titik data yang lebih banyak dibandingkan dengan kluster 2. Namun, terdapat beberapa *outlier* yang terdeteksi, di mana beberapa titik data terlalu jauh dari titik data lainnya. *DBSCAN* menganggap titik-titik ini sebagai outlier dan memberikan label -1 pada titik-titik tersebut.

Untuk mengevaluasi hasil klustering ini lebih lanjut, perlu dianalisis kelebihan dan kelemahan dari kluster yang terbentuk serta implikasinya terhadap segmentasi pelanggan. Kelebihan dari hasil klustering DBSCAN adalah kemampuannya untuk mendeteksi kluster dengan kepadatan yang bervariasi dan mengidentifikasi outlier secara efektif. Kluster 1, yang memiliki jumlah titik data lebih banyak, kemungkinan besar mewakili kelompok pelanggan dengan karakteristik yang dominan, seperti pengguna kartu kredit reguler. Sementara itu, kluster 2 yang lebih kecil mungkin mewakili segmen pelanggan dengan perilaku yang lebih spesifik atau unik. Outlier yang terdeteksi oleh DBSCAN (diberi label -1) dapat menjadi fokus untuk analisis lebih lanjut, seperti deteksi penipuan atau identifikasi pelanggan dengan kebutuhan khusus. Namun, kelemahan dari hasil klustering ini adalah kesulitan dalam menentukan parameter Epsilon dan MinPoints yang optimal, terutama jika data memiliki kepadatan yang tidak seragam. Selain itu, DBSCAN mungkin kurang efektif jika data memiliki kluster dengan kepadatan yang sangat berbeda, karena parameter yang sama mungkin tidak cocok untuk semua kluster.

3.2.3 Hierarchical Clustering (Agglomerative)

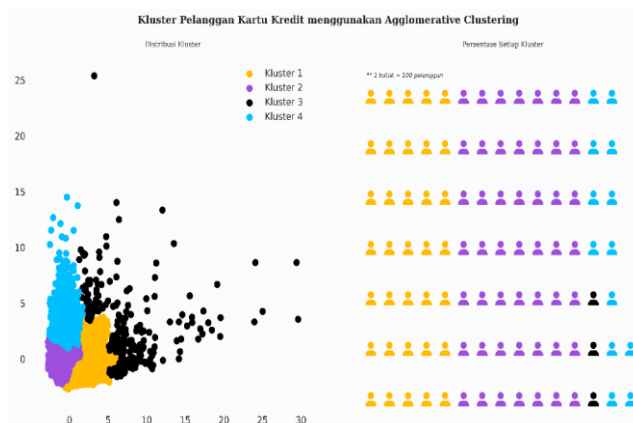
Pada bagian ini, algoritma *Hierarchical Clustering* diterapkan untuk mengelompokkan data pelanggan kartu kredit. Metode ini berfungsi dengan mengelompokkan data ke dalam struktur pohon kluster. Proses dimulai dengan menganggap setiap titik data sebagai kluster terpisah. Algoritma ini kemudian secara berulang mengidentifikasi dua kluster yang paling dekat satu sama lain dan menggabungkan dua kluster yang paling dapat dibandingkan hingga semua kluster bergabung menjadi satu.

Dendrogram digunakan untuk memvisualisasikan sejarah pengelompokan dan menentukan jumlah kluster yang optimal. Dengan menggunakan *dendrogram* yang dihasilkan, kami dapat menentukan jarak vertikal terbesar yang tidak memotong kluster lainnya. Garis horizontal ditarik pada kedua ekstrem, dan jumlah kluster optimal ditentukan berdasarkan jumlah garis vertikal yang melewati garis horizontal tersebut. Dalam analisis ini, jumlah kluster optimal yang diperoleh adalah 4, seperti yang ditunjukkan pada Gambar 9.



Gambar 9. Dendrogram dan Skor Calinski-Harabasz Elbow

Berdasarkan jarak *Euclidean* dalam *dendrogram*, dapat disimpulkan bahwa jumlah kluster yang terbentuk adalah empat, karena garis vertikal tertinggi terletak pada cabang pertama di sebelah kiri gambar. Selain itu, *skor Calinski-Harabasz* juga mendukung hasil ini dengan menunjukkan bahwa jumlah kluster optimal adalah 4. Setelah menentukan jumlah kluster, langkah selanjutnya adalah menerapkan algoritma *Agglomerative Clustering* dan memvisualisasikan serta mengevaluasi kluster yang terbentuk, seperti yang ditunjukkan pada Gambar 10.



Gambar 10. Hasil Kluster Pelanggan menggunakan Hierarchical Clustering (Agglomerative)

Hasil implementasi menunjukkan bahwa terdapat 4 kluster yang terbentuk. Dari keempat kluster tersebut, kluster 2 memiliki jumlah titik data terbanyak, diikuti oleh kluster 1. Ketika dibandingkan dengan hasil klustering menggunakan *K-Means*, kluster 2 dalam hasil hierarkis menunjukkan persentase yang lebih signifikan. Selain itu,

algoritma klustering hierarkis menganggap *outlier* sebagai bagian dari kluster 3, yang memberikan wawasan tambahan tentang bagaimana algoritma ini menangani data yang tidak biasa.

Untuk memahami lebih dalam implikasi dari hasil klustering ini, penting untuk menganalisis kelebihan dan kelemahan dari setiap kluster yang terbentuk. Kelebihan dari hasil klustering Hierarchical Clustering adalah kemampuannya untuk menghasilkan struktur kluster yang hierarkis, yang memungkinkan analisis lebih mendalam tentang hubungan antara kluster. Kluster 2, yang memiliki jumlah titik data terbanyak, kemungkinan besar mewakili segmen pelanggan utama, seperti pengguna kartu kredit dengan tingkat pengeluaran menengah. Kluster 1 dan 3 mungkin mewakili sub-segmen yang lebih spesifik, seperti pelanggan dengan pengeluaran rendah atau tinggi. *Outlier* yang termasuk dalam kluster 3 dapat memberikan wawasan tambahan tentang bagaimana algoritma ini menangani data yang tidak biasa, yang mungkin berguna untuk analisis risiko atau penawaran produk khusus. Namun, kelemahan dari hasil klustering ini adalah kompleksitas komputasinya yang tinggi, terutama untuk dataset yang besar. Selain itu, Hierarchical Clustering bersifat deterministik, artinya sekali kluster terbentuk, tidak ada cara untuk membatalkan penggabungan yang telah dilakukan, yang dapat menjadi masalah jika terdapat kesalahan dalam proses pengelompokan.

3.3 Evaluasi Model

Pada bagian ini, evaluasi kualitas hasil klustering dari algoritma yang telah diterapkan akan dilakukan. Evaluasi ini akan membandingkan hasil klustering dari masing-masing algoritma menggunakan tiga metrik utama: *indeks Davies-Bouldin*, *skor silhouette*, dan *indeks Calinski-Harabasz*. Hasil evaluasi ini akan disajikan dalam format tabel (Tabel 2).

Tabel 2. Hasil Evaluasi Model

Model	Davies-Bouldin Index	Silhouette Score	Calinski-Harabasz Index
K-Means	0.800000	0.408000	5823.343000
Hierarchical Clustering	0.855000	0.388000	4936.960000
DBSCAN	1.287000	0.803000	685.302000

Tabel 2 menunjukkan bahwa algoritma *K-Means* memiliki *indeks Davies-Bouldin* terendah dibandingkan dengan dua algoritma lainnya, sehingga dapat disimpulkan bahwa *K-Means* memiliki kualitas klustering yang baik dibandingkan dengan kedua algoritma lainnya. Namun, berdasarkan *skor silhouette*, *K-Means* memiliki *skor silhouette* tertinggi kedua, yang menunjukkan adanya beberapa kluster yang tumpang tindih yang terbentuk menggunakan algoritma ini.

Selanjutnya, klustering menggunakan algoritma *Hierarchical Clustering* menunjukkan hasil kualitas klustering yang serupa dengan *K-Means*. *Indeks Davies-Bouldin* sedikit lebih tinggi, dan skor *silhouette* sedikit lebih rendah dibandingkan dengan *K-Means*. Akhirnya, klustering menggunakan *DBSCAN* menunjukkan *indeks Davies-Bouldin* terburuk tetapi memiliki *skor silhouette* terbaik dibandingkan dengan algoritma lainnya.

Dari hasil *indeks Calinski-Harabasz*, terlihat bahwa *K-Means* memiliki indeks tertinggi dibandingkan dengan algoritma lainnya. Ini menunjukkan bahwa *K-Means* berkinerja lebih baik dan lebih padat dibandingkan dengan algoritma lainnya. Dapat disimpulkan bahwa *K-Means* memiliki kualitas klustering terbaik di antara ketiga algoritma tersebut, berdasarkan nilai *indeks Davies-Bouldin* terendah dan kluster yang sedikit lebih baik dibandingkan dengan klustering hierarkis. Langkah selanjutnya adalah melakukan profiling untuk mengidentifikasi karakteristik masing-masing kluster.

3.4 Analisis dan Profiling

Pada bagian ini, dilakukan profiling kluster untuk mengidentifikasi karakteristik dari kluster yang telah dibentuk oleh algoritma *K-Means*, seperti yang ditunjukkan pada Gambar 11.

Nama Kolom	Metode	Kluster 1	Kluster 2	Kluster 3	Kluster 4	Keseluruhan
Saldo	mean	814.406180	1028.913422	3404.277246	4529.658912	1564.474828
Frekuensi_Saldo	mean	0.917771	0.800633	0.988518	0.964222	0.877271
Pembelian	mean	1224.752724	219.106636	6866.239428	465.488779	1003.204834
Pembelian_Sekali	mean	615.489080	155.453341	4495.247515	302.444841	592.437371
Pembelian_Cicilan	mean	609.466137	63.984949	2372.175345	163.144565	411.067645
Uang_Muka	mean	147.839231	625.069377	771.262149	4432.805957	978.871112
Frekuensi_Pembelian	mean	0.858322	0.185533	0.954802	0.267226	0.490351
Frekuensi_Pembelian_Sekali	mean	0.299198	0.073954	0.727673	0.130512	0.202458
Frekuensi_Pembelian_Cicilan	mean	0.671196	0.106485	0.808810	0.166994	0.364437
Frekuensi_Uang_Muka	mean	0.030657	0.124503	0.084630	0.472418	0.135144
Transaksi_Uang_Muka	mean	0.567949	2.306842	2.351085	13.854353	3.248827
Transaksi_Pembelian	mean	21.108839	2.889660	82.684418	6.744508	14.709832
Batas_Kredit	mean	4232.252030	3117.619170	9521.597633	7484.706709	4494.293646
Pembayaran	mean	1318.695203	862.288423	6687.644858	3565.861596	1733.143852
Pembayaran_Minimum	mean	595.536472	596.244012	1823.365502	2076.172933	868.716633
Pembayaran_Penuh	mean	0.282278	0.064067	0.285741	0.034927	0.153715
Tenur	mean	11.656043	11.364969	11.960552	11.441009	11.517318

Gambar 11. Hasil Ringkasan Kluster

Berdasarkan Gambar 11, dapat disimpulkan bahwa setiap cluster memiliki karakteristik sebagai berikut:

- a. Kluster 1 (Pengguna Pembayar Penuh): Pelanggan dalam kluster ini adalah pengguna aktif kartu kredit bank. Hal ini terlihat dari frekuensi saldo yang sering berubah dan jumlah saldo yang cukup tinggi dibandingkan dengan kluster lainnya. Kluster ini juga memiliki nilai rata-rata yang lebih tinggi dalam beberapa aspek dibandingkan kluster lainnya. Pelanggan di kluster ini aktif menggunakan kartu kredit untuk mempermudah transaksi dan cicilan, dengan frekuensi transaksi dan cash advance yang lebih tinggi. Tenor yang relatif tinggi menunjukkan bahwa penilaian kredit di kluster ini sangat baik.
- b. Kluster 2 (Pengguna Pemula/Mahasiswa): Berbeda dengan kluster 1, pelanggan di kluster ini jarang atau hampir tidak pernah menggunakan kartu kredit untuk transaksi dan cicilan. Hal ini disebabkan oleh saldo yang relatif kecil, frekuensi saldo yang jarang berubah, dan cicilan yang sangat rendah. Batas kredit yang rendah juga menunjukkan bahwa pelanggan di kluster ini jarang menggunakan kartu kredit untuk transaksi kredit. Dapat diasumsikan bahwa pelanggan di kluster ini adalah mahasiswa atau pengguna baru yang menggunakan kartu kredit di bank ini.
- c. Kluster 3 (Pengguna Cicilan): Dalam kluster ini, pelanggan menggunakan kartu kredit khusus untuk tujuan cicilan. Hal ini terlihat dari tingkat transaksi menggunakan cicilan yang relatif tinggi. Pelanggan di kluster ini sering melakukan transaksi dengan jumlah yang sangat besar per transaksi, sementara frekuensi dan transaksi cash advance sangat kecil. Pelanggan di kluster ini sangat jarang melakukan pembayaran dan *cash advance*, sehingga dapat disimpulkan bahwa mereka sangat cocok untuk kartu kredit yang khusus untuk kebutuhan cicilan.
- d. Kluster 4 (Pengguna Penarikan Tunai): Pelanggan dalam kluster ini memiliki saldo tinggi, frekuensi saldo yang selalu berubah, dan frekuensi cash advance yang tinggi. Selain itu, pelanggan di kluster ini memiliki suku bunga terendah dibandingkan dengan kluster lainnya dan memiliki batas kredit serta pembayaran tertinggi kedua dari empat kluster. Namun, pengguna kartu kredit di kluster ini jarang melakukan cicilan atau pembelian satu kali dan memiliki tenor ketiga tertinggi dari empat kluster. Dengan demikian, dapat disimpulkan bahwa pelanggan di kluster ini hanya menggunakan kartu kredit untuk kebutuhan penarikan uang atau *cash advance*.

4. KESIMPULAN

Dalam penelitian ini, telah dilakukan analisis klustering untuk mengidentifikasi karakteristik pelanggan kartu kredit dan merumuskan strategi pemasaran yang sesuai berdasarkan hasil *profiling*. Hasil analisis menunjukkan bahwa pelanggan dalam kluster 1, yang merupakan pengguna aktif dengan tenure dan batas kredit tertinggi, dapat menjadi target utama dalam pemasaran kartu kredit. Dengan menawarkan manfaat dan *reward*, bank dapat mendorong penggunaan kartu kredit yang lebih sering, sehingga meningkatkan profitabilitas. Sementara itu, pelanggan di kluster 3, yang lebih cenderung menggunakan kartu kredit untuk cicilan, dapat dijadikan fokus untuk program cicilan dengan bunga rendah atau tanpa bunga, yang dapat menarik minat mereka untuk memanfaatkan kartu kredit secara optimal. Kluster 2, yang terdiri dari pemula atau mahasiswa, memerlukan pendekatan khusus dengan menawarkan kartu kredit yang dirancang untuk membantu mereka membangun profil kredit dan belajar menggunakan kartu secara bertanggung jawab. Terakhir, pelanggan di kluster 4, yang sering melakukan penarikan tunai, dapat diberikan penawaran kartu kredit dengan biaya rendah dan manfaat tambahan. Meskipun penelitian ini memberikan wawasan yang berharga, terdapat beberapa keterbatasan yang perlu diperhatikan, seperti ukuran sampel yang mungkin tidak mencakup seluruh populasi pelanggan dan variabel yang tidak diperhitungkan dalam analisis. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas cakupan data dengan mempertimbangkan aspek demografis, psikografis, serta faktor eksternal seperti kondisi ekonomi dan kebijakan perbankan. Dengan memperkaya variabel analisis, hasil penelitian ini dapat digunakan sebagai dasar untuk pengembangan strategi pemasaran yang lebih efektif dan relevan di masa depan.

REFERENCES

- [1] I. M. Karo Karo, "Segmentation of Credit Card Customers Based on Their Credit Card Usage Behavior using The K-Means Algorithm," *J. Softw. Eng. Inf. Commun. Technol.*, vol. 2, no. 2, pp. 55–64, 2022, doi: 10.17509/seict.v2i2.40220.
- [2] C. A. Robb, "Financial Knowledge and Credit Card Behavior of College Students," *J. Fam. Econ. Issues*, vol. 32, no. 4, pp. 690–698, 2011, doi: 10.1007/s10834-011-9259-y.
- [3] A. Fathurozi and T. Ginanjar Laksana, "Enhancing Promotional Strategy Mapping Using the K-Means Clustering Algorithm to Raise Sales," *is Best Account. Inf. Syst. Inf. Technol. Bus. Enterp. this is link OJS us*, vol. 8, no. 2, pp. 121–135, 2024, doi: 10.34010/aisthebest.v8i2.11597.
- [4] Y. Wicaksono, "Segmentasi Pelanggan Bisnis Dengan Multi Kriteria Menggunakan K-Means," *Indones. J. Bus. Intell.*, vol. 1, no. 2, p. 45, 2019, doi: 10.21927/ijubi.v1i2.872.
- [5] M. F. Fadhillah, A. Lovely, A. Suyoso, and I. Puspitasari, "Customer Segmentation with Clustering Algorithm Based on Recency, Frequency, and Monetary (RFM) Attributes Segmentasi Pelanggan dengan Algoritma Clustering Berdasarkan Atribut Recency, Frequency dan Monetary (RFM)," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. January, pp. 48–56, 2025.
- [6] A. S. P. Paramita, "Comparison of K-Means and DBSCAN Algorithms for Customer Segmentation in E-commerce," *J. Digit. Mark. Digit. Curr.*, vol. 1, no. 1, pp. 29–43, 2024, doi: 10.47738/jdmdc.v1i1.3.
- [7] A. Abdulhafedh, "Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation," *J. City Dev.*, vol. 3, no. 1, pp. 12–30, 2021, doi: 10.12691/jcd-3-1-3.



- [8] H. Mukhtar, I. D. Pramaditya, W. S. Weisdianto, and S. H. Putra, “Algoritma K-Means Untuk Pengelompokan Perilaku Customer,” *J. Softw. Eng. Inf. Syst.*, vol. 4, no. 2, pp. 96–101, 2024.
- [9] Baiq Nikum Yuliasih, H. Herman, and S. Sunardi, “K-Means Clustering Method For Customer Segmentation Based On Potential Purchases,” *J. ELTIKOM*, vol. 8, no. 1, pp. 83–90, 2024, doi: 10.31961/eltikom.v8i1.1137.
- [10] 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.
- [11] L. Abednego, C. E. Nugraheni, and A. Salsabina, “Customer Segmentation: Transformation from Data to Marketing Strategy,” *Conf. Ser.*, vol. 4, no. 1, pp. 139–152, 2023, doi: 10.34306/conferenceseries.v4i1.645.
- [12] Y. Qiu and J. Wang, “A Machine Learning Approach to Credit Card Customer Segmentation for Economic Stability,” *Proc. 4th Int. Conf. Econ. Manag. Big Data Appl. ICEMBDA 2023, Oct. 27–29, 2023, Tianjin, China*, 2024, doi: 10.4108/eai.27-10-2023.2342007.
- [13] W. Rohalidyawati, R. Rahmawati, and M. Mustafid, “Segmentasi Pelanggan E-Money Dengan Menggunakan Algoritma Dbscan (Density Based Spatial Clustering Applications With Noise) Di Provinsi Dki Jakarta,” *J. Gaussian*, vol. 9, no. 2, pp. 162–169, 2020, doi: 10.14710/j.gauss.v9i2.27818.
- [14] F. M. Pranata, S. H. Wijoyo, and N. Y. Setiawan, “Analisis Performa Algoritma K-Means dan DBSCAN Dalam Segmentasi Pelanggan Dengan Pendekatan Model RFM,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 1, pp. 2548–964, 2017, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/13962>
- [15] M. Sarkar, A. R. Puja, and F. R. Chowdhury, “Optimizing Marketing Strategies with RFM Method and K-Means Clustering-Based AI Customer Segmentation Analysis,” *J. Bus. Manag. Stud.*, vol. 6, no. 2, pp. 54–60, 2024, doi: 10.32996/jbms.2024.6.2.5.
- [16] H. H. Zhao, X. C. Luo, R. Ma, and X. Lu, “An Extended Regularized K-Means Clustering Approach for High-Dimensional Customer Segmentation with Correlated Variables,” *IEEE Access*, vol. 9, pp. 48405–48412, 2021, doi: 10.1109/ACCESS.2021.3067499.
- [17] S. Lade, “CREDIT CARD-SEGMENTATION,” *SSRN*, 2024, doi: <https://dx.doi.org/10.2139/ssrn.4939839>.
- [18] M. A. Fitriani and D. C. Febrianto, “Data Mining for Potential Customer,” *JUITA J. Inform.*, vol. 9, no. 1, pp. 25–32, 2021, doi: 10.30595/juita.v9i1.7983.
- [19] E. L. Cahapin, B. A. Malabag, C. S. Santiago, J. L. Reyes, G. S. Legaspi, and K. L. Adrales, “Clustering of students admission data using k-means, hierarchical, and DBSCAN algorithms,” *Bull. Electr. Eng. Informatics*, vol. 12, no. 6, pp. 3647–3656, 2023, doi: 10.11591/eei.v12i6.4849.
- [20] D. Hendarsyah, “Analisis Perilaku Konsumen Dan Keamanan Kartu Kredit Perbankan,” *JPS (Jurnal Perbank. Syariah)*, vol. 1, no. 1, pp. 85–96, 2020, doi: 10.46367/jps.v1i1.204.
- [21] V. Mihova and V. Pavlov, “A customer segmentation approach in commercial banks,” *AIP Conf. Proc.*, vol. 2025, no. 1, p. 30003, Oct. 2018, doi: 10.1063/1.5064881.
- [22] F. P. Rachman, H. Santoso, and A. Djajadi, “Machine Learning Mini Batch K-means and Business Intelligence Utilization for Credit Card Customer Segmentation,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 10, pp. 218–227, 2021, doi: 10.14569/IJACSA.2021.0121024.