

Mengungkap Pola Konsumen dengan Model Prediktif dan Clustering pada Data Kredit Pelanggan

Yunita Dwi Putri ¹, Fenni Aprilia ², Rasti Aulia Anggraini ³

^{1,2,3} Politeknik Negeri Lampung, Teknologi Rekayasa Internet

INFORMASI ARTIKEL

Diterima 27 Juni 2025
Direvisi 10 Juli 2025
Diterbitkan 11 Juli 2025

Kata kunci:

Konsumen;
Pola;
Kredit;
Clustering;
Prediksi

ABSTRAK

Penelitian ini dilatarbelakangi oleh tantangan yang dihadapi perusahaan dalam mengelola volume data pelanggan yang besar untuk mendukung pengambilan keputusan strategis. Data keuangan pelanggan, seperti saldo rekening dan batas kredit, memiliki potensi besar untuk mengungkap pola perilaku yang relevan tetapi memerlukan analisis sistematis. Penelitian ini menggunakan pendekatan machine learning untuk mengelompokkan pelanggan berdasarkan karakteristik keuangan mereka dan memprediksi perilaku pembayaran. Model clustering diterapkan dengan algoritma K-Means untuk membagi pelanggan ke dalam tiga segmen utama berdasarkan saldo, pembelian, dan pembayaran mereka. Hasil clustering divisualisasikan menggunakan Principal Component Analysis (PCA) untuk memahami distribusi antar segmen. Selain itu, model prediktif berbasis regresi linier digunakan untuk memprediksi pembayaran pelanggan berdasarkan atribut keuangan utama. Metode penelitian mencakup pengumpulan data kartu kredit pelanggan, persiapan data melalui normalisasi dan pengisian nilai hilang, analisis kluster menggunakan K-Means, serta evaluasi prediksi dengan metrik Mean Squared Error dan R-squared. Penelitian ini menggunakan dataset yang berisi informasi keuangan pelanggan kartu kredit dengan 18 atribut dan 8.950 entri. Hasil penelitian menunjukkan bahwa algoritma K-Means berhasil mengidentifikasi tiga kluster dengan karakteristik keuangan yang berbeda. Kluster pertama mencakup pelanggan dengan saldo sedang dan frekuensi pembelian tinggi. Kluster kedua terdiri dari pelanggan dengan saldo rendah hingga sedang dan aktivitas keuangan yang stabil. Kluster ketiga adalah pelanggan dengan saldo dan batas kredit tinggi tetapi aktivitas pembelian rendah. Model regresi linier berhasil memprediksi pembayaran pelanggan dengan akurasi yang memadai.

Unveiling Consumer Patterns Through Predictive Modeling and Clustering on Customer Credit Data

ARTICLE INFO

Received June 27, 2025
Revised July 10, 2025
Published July 11, 2025

Keyword:

Consumers;
Patterns;
Credit;
Clustering;
Prediction

ABSTRACT

This research is motivated by the challenges faced by companies in managing large volumes of customer data to support strategic decision-making. Customer financial data, such as account balances and credit limits, hold great potential for uncovering relevant behavioral patterns but require systematic analysis. This study employs a machine learning approach to cluster customers based on their financial characteristics and to predict payment behavior. The clustering model is implemented using the K-Means algorithm to segment customers into three main groups based on their balances, purchases, and payments. The clustering results are visualized using Principal Component Analysis (PCA) to better understand the distribution among segments. In addition, a predictive model using linear regression is applied to forecast customer payments based on key financial attributes. The research methodology includes collecting credit card customer data, preparing the data through normalization and handling missing values, conducting cluster analysis using K-Means, and evaluating the prediction model using Mean Squared Error (MSE) and R-squared metrics. The dataset used contains financial information from 8,950 entries and 18 attributes of credit card customers. The results show that the K-Means algorithm successfully identified three clusters with distinct financial characteristics. The first cluster consists of customers with moderate balances and high purchase frequency. The second cluster includes customers with low to moderate balances and stable financial activity. The third cluster comprises customers with high balances and credit limits but low purchasing activity. The linear regression model was able to predict customer payments with satisfactory accuracy.

This work is licensed under a [Creative Commons Attribution 4.0](#)



Corresponding Author:

Yunita Dwi Putri, Teknologi Rekayasa Internet
Email: yunitadwiputri88@gmail.com

1. PENDAHULUAN

Data pelanggan merupakan salah satu aset penting dalam strategi bisnis modern, khususnya dalam industri jasa keuangan seperti perbankan dan kartu kredit. Melalui data ini, perusahaan

dapat memahami perilaku konsumsi, preferensi, hingga risiko keuangan pelanggan[1]. Di era digital, volume data yang dihasilkan dari aktivitas pelanggan terus meningkat, mencakup informasi transaksi, pola pembayaran, hingga penggunaan kredit. Namun, besarnya volume data ini menjadi tantangan dalam mengidentifikasi pola yang signifikan dan relevan untuk mendukung pengambilan keputusan strategis. Oleh karena itu, diperlukan pendekatan analitis yang sistematis untuk mengolah data ini menjadi informasi yang dapat diimplementasikan secara efektif[2].

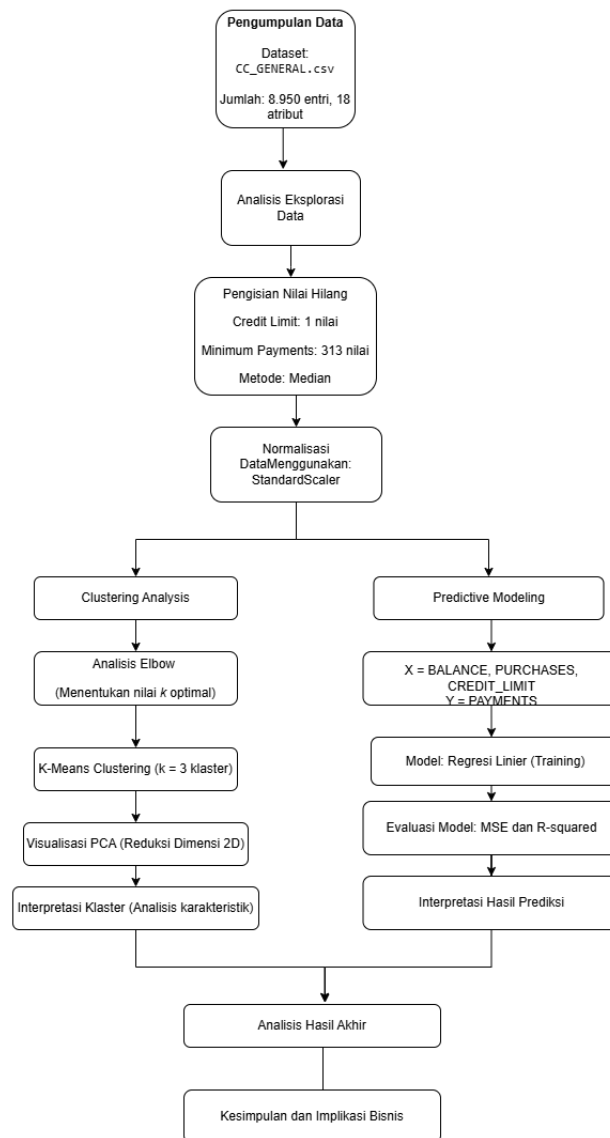
Dalam penelitian ini, teknik machine learning digunakan untuk mengungkap pola dan kecenderungan pelanggan berdasarkan data keuangan mereka. Clustering diterapkan untuk mengelompokkan pelanggan berdasarkan kesamaan karakteristik keuangan, seperti saldo rekening, batas kredit, dan pembayaran. Dengan menggunakan algoritma K-Means, pelanggan dikelompokkan ke dalam beberapa segmen yang dapat membantu perusahaan memahami profil dan kebutuhan pelanggan secara lebih mendalam. Selanjutnya, analisis komponen utama (Principal Component Analysis) digunakan untuk memvisualisasikan hasil clustering dalam ruang dua dimensi, sehingga pola antar kelompok dapat terlihat secara jelas[3].

Selain itu, predictive modeling menggunakan regresi linier diterapkan untuk memprediksi pembayaran pelanggan berdasarkan atribut utama, seperti saldo dan batas kredit. Model ini dirancang untuk membantu perusahaan dalam memperkirakan pembayaran pelanggan di masa depan, yang dapat digunakan sebagai dasar untuk pengelolaan risiko kredit dan strategi pemasaran.

Pendekatan ini tidak hanya memberikan wawasan yang lebih mendalam tentang perilaku pelanggan tetapi juga mendukung perencanaan strategi bisnis yang lebih personal dan berbasis data. Dengan memahami pola pelanggan secara lebih baik, perusahaan dapat merancang program loyalitas, meningkatkan layanan, dan mengidentifikasi peluang untuk mengoptimalkan profitabilitas. Penelitian ini bertujuan untuk menunjukkan bagaimana kombinasi teknik clustering dan predictive modeling dapat diaplikasikan pada data keuangan pelanggan untuk menghasilkan insight yang bernilai dalam mendukung pengambilan keputusan berbasis fakta[4].

2. METODE

Metode penelitian ini dirancang secara sistematis untuk menganalisis pola konsumen menggunakan model prediktif dan clustering pada data pelanggan kartu kredit. Penelitian dilakukan melalui beberapa tahapan yang saling berkaitan, mulai dari pengumpulan data, pengolahan data, hingga analisis menggunakan teknik machine learning. Setiap tahapan dirancang untuk menghasilkan informasi yang relevan dan dapat diinterpretasikan untuk pengambilan keputusan strategis[5].



Gambar 1. Diagram Alur Metodologi Penelitian

Gambar 1 menunjukkan tahapan sistematis yang dilakukan dalam penelitian ini. Dimulai dari pengumpulan dataset CC_GENERAL.csv yang berisi 8.950 entri pelanggan kartu kredit dengan 18 atribut. Tahap persiapan data mencakup penanganan missing values dan normalisasi menggunakan StandardScaler untuk memastikan kualitas data yang optimal. Proses clustering menggunakan algoritma K-Means dengan penentuan jumlah klaster optimal melalui metode elbow, diikuti dengan visualisasi menggunakan Principal Component Analysis (PCA). Secara paralel, model prediktif regresi linier dikembangkan untuk memprediksi pembayaran pelanggan. Seluruh proses diakhiri dengan analisis dan interpretasi hasil yang mendalam untuk mendukung pengambilan keputusan bisnis yang berbasis data.

2.1. Pengumpulan Data

Tahap pertama dalam penelitian ini adalah pengumpulan data pelanggan kartu kredit yang diperoleh dari platform Kaggle. Dataset yang digunakan memiliki nama file CC_GENERAL.csv, berisi 18 atribut dan 8.950 entri yang mencakup berbagai informasi keuangan pelanggan. Atribut utama dalam dataset ini meliputi balance (saldo rekening), purchases (frekuensi pembelian), credit limit (batas kredit), payments (pembayaran), dan atribut lain yang mencerminkan aktivitas

keuangan pelanggan. Dataset ini dipilih karena menyediakan informasi yang kaya dan relevan untuk menganalisis pola perilaku keuangan pelanggan[6].

2.2. Persiapan Data

Tahap berikutnya adalah persiapan data, yang mencakup analisis struktur dataset dan pengolahan nilai yang hilang. Dataset diperiksa untuk mendeteksi data kosong, tidak konsisten, atau outlier. Untuk atribut numerik dengan nilai yang hilang, seperti pada kolom credit limit (1 nilai hilang) dan minimum payments (313 nilai hilang), pengisian dilakukan menggunakan nilai median. Metode ini dipilih untuk mengurangi bias akibat data ekstrem[7].

Selanjutnya, atribut numerik dalam dataset dinormalisasi menggunakan StandardScaler. Normalisasi ini bertujuan untuk menyelaraskan skala antar atribut agar analisis tidak terdistorsi oleh perbedaan skala, khususnya dalam proses clustering yang sensitif terhadap magnitudo data.

2.3. Clustering dengan K-Means

Untuk mengelompokkan pelanggan berdasarkan kesamaan pola keuangan, algoritma K-Means digunakan. Empat atribut utama yang digunakan dalam proses clustering adalah balance, purchases, credit limit, dan payments. Sebelum penerapan K-Means, dilakukan analisis elbow untuk menentukan jumlah kluster optimal[8].

Metode elbow mengevaluasi nilai inertia (jumlah kuadrat kesalahan dalam kluster) untuk berbagai jumlah kluster, dan titik optimal ditentukan dari perubahan nilai inertia yang signifikan. Berdasarkan analisis ini, tiga kluster ($k=3$) dipilih sebagai jumlah kluster yang optimal. Setelah itu, algoritma K-Means diterapkan dengan parameter optimal untuk mengelompokkan pelanggan, dan hasil clustering ditambahkan sebagai kolom baru di dataset.

Visualisasi hasil clustering dilakukan dengan Principal Component Analysis (PCA). PCA mereduksi dimensi data menjadi dua komponen utama, yang memungkinkan visualisasi kluster dalam ruang dua dimensi untuk mempermudah interpretasi perbedaan antar kluster.

2.4. Model Prediktif dengan Regresi Linier

Untuk memprediksi pembayaran pelanggan (PAYMENTS), digunakan model regresi linier dengan atribut utama sebagai variabel independen, yaitu BALANCE, PURCHASES, dan CREDIT_LIMIT. Model dilatih menggunakan seluruh dataset, dan metrik evaluasi seperti Mean Squared Error (MSE) dan R-squared (R^2) dihitung untuk mengevaluasi kinerja model prediktif. Hasil prediksi kemudian dibandingkan dengan data aktual untuk memahami kemampuan model dalam menjelaskan variasi data.

2.5. Model Prediktif dengan Regresi Linier Analisis dan Interpretasi Hasil

Hasil clustering dianalisis untuk memahami karakteristik utama dari masing-masing kluster. Setiap kluster dievaluasi berdasarkan distribusi atribut yang mendasari pembentukan kluster tersebut. Kluster yang dihasilkan diinterpretasikan sebagai segmen pelanggan dengan pola keuangan yang serupa, seperti kelompok pelanggan dengan saldo tinggi, pembelian rendah, atau pembayaran yang tidak teratur[9].

Selain itu, hasil prediksi model regresi linier diinterpretasikan untuk mengevaluasi efektivitas model dalam memprediksi pembayaran pelanggan. Pola hubungan antara atribut utama dan pembayaran dianalisis untuk mengidentifikasi faktor-faktor yang paling memengaruhi perilaku pembayaran pelanggan.

3. HASIL DAN PEMBAHASAN

Pada bagian ini dijelaskan hasil penelitian mencakup analisis dataset pelanggan kartu kredit untuk mengidentifikasi pola konsumsi menggunakan metode clustering serta memprediksi pembayaran pelanggan dengan model prediktif.

3.1. Eksplorasi Dataset

Eksplorasi dataset merupakan langkah awal untuk memahami karakteristik data yang digunakan dalam penelitian ini. Dataset terdiri dari 8.950 entri dan 18 atribut, yang mencakup informasi seperti saldo rekening (balance), frekuensi pembelian (purchases), batas kredit (credit

limit), pembayaran (payments), dan pembayaran minimum (minimum payments). Selama eksplorasi, ditemukan adanya nilai yang hilang pada kolom credit limit (1 nilai hilang) dan minimum payments (313 nilai hilang)[10].

Pengisian nilai hilang dilakukan menggunakan metode median untuk menjaga stabilitas analisis tanpa terpengaruh oleh outlier. Selain itu, normalisasi atribut numerik dengan StandardScaler memastikan data berada dalam skala yang sama, sehingga setiap atribut memiliki bobot yang setara dalam analisis selanjutnya.

3.2. Clustering Pelanggan

Analisis clustering dilakukan menggunakan algoritma K-Means. Metode Elbow digunakan untuk menentukan jumlah kluster ideal, yang menghasilkan tiga kluster utama. Kluster pertama mencakup 1.968 pelanggan dengan saldo rekening yang berada pada tingkat sedang. Frekuensi pembelian pelanggan dalam kluster ini relatif tinggi, dengan kecenderungan dominasi pembelian melalui angsuran. Kluster kedua, yang merupakan kluster terbesar, terdiri dari 6.833 pelanggan dengan karakteristik keuangan yang stabil. Pelanggan dalam kluster ini memiliki saldo yang cenderung rendah hingga sedang, serta aktivitas pembelian yang merata. Sementara itu, kluster ketiga mencakup 149 pelanggan yang memiliki saldo tinggi dan batas kredit yang sangat besar, tetapi dengan frekuensi pembelian yang lebih rendah[11].

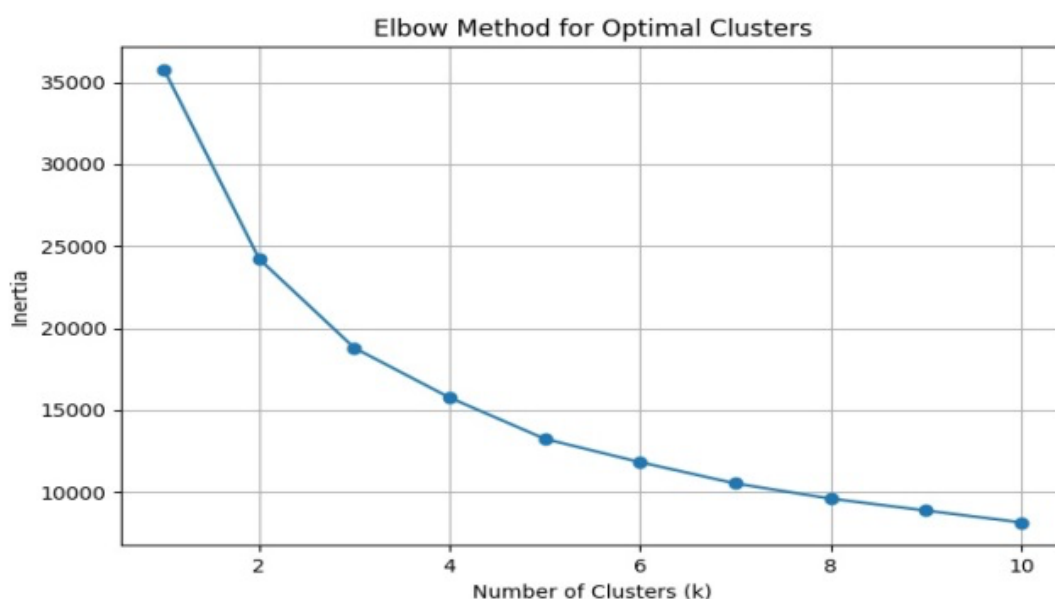
Tabel 1. Karakteristik Kluster Pelanggan

Kluster	Jumlah Pelanggan	Persentase	Karakteristik Saldo	Karakteristik Pembelian	Karakteristik Kredit	Profil Utama
Kluster 0	1.968	21.98%	Saldo Sengah	Frekuensi Tinggi	Batas Kredit Moderat	Pelanggan Aktif dengan pembelian angsuran dominan
Kluster 1	6.833	76.35%	Saldo Rendah-Sengah	Aktivitas Merata	Batas Kredit Standar	Pelanggan Mainstream dengan aktivitas stabil
Kluster 2	149	1.67%	Saldo Tinggi	Frekuensi Rendah	Batas Kredit Tinggi	Pelanggan Premium dengan kapasitas finansial besar
Total	8.95	100%	-	-	-	-

Hasil analisis clustering menggunakan algoritma K-Means menghasilkan tiga kluster dengan karakteristik yang berbeda, sebagaimana ditunjukkan dalam Tabel 1. Kluster pertama mencakup 1.968 pelanggan dengan saldo rekening yang berada pada tingkat sedang. Frekuensi pembelian pelanggan dalam kluster ini relatif tinggi, dengan kecenderungan dominasi pembelian melalui angsuran. Kluster kedua, yang merupakan kluster terbesar, terdiri dari 6.833 pelanggan dengan karakteristik keuangan yang stabil. Pelanggan dalam kluster ini memiliki saldo yang cenderung rendah hingga sedang, serta aktivitas pembelian yang merata. Sementara itu, kluster ketiga mencakup 149 pelanggan yang memiliki saldo tinggi dan batas kredit yang sangat besar, tetapi dengan frekuensi pembelian yang lebih rendah.

Berdasarkan Tabel 1, dapat diamati bahwa distribusi pelanggan menunjukkan pola yang menarik. Kluster 1 mendominasi dengan 76,35% dari total pelanggan, mencerminkan bahwa mayoritas pengguna kartu kredit termasuk dalam kategori pelanggan mainstream dengan aktivitas keuangan yang stabil. Kluster 0 dengan 21,98% pelanggan menunjukkan segmen yang aktif dalam bertransaksi meskipun dengan saldo yang moderat. Sementara itu, Kluster 2 meski hanya mencakup

1,67% pelanggan, memiliki nilai strategis tinggi karena merepresentasikan segmen premium dengan kapasitas finansial yang besar



Gambar 2. Visualisasi Elbow

Gambar 2 menunjukkan hasil analisis metode elbow yang digunakan untuk menentukan jumlah kluster optimal dalam penerapan algoritma K-Means. Grafik pada gambar ini memperlihatkan hubungan antara jumlah kluster (k) dengan nilai inertia, yaitu ukuran seberapa baik data dikelompokkan dalam kluster. Penurunan nilai inertia terlihat signifikan hingga jumlah kluster $k=3$, setelah itu penurunan menjadi lebih lambat. Titik "sudut" pada grafik, yang dikenal sebagai elbow point, berada pada $k=3$. Oleh karena itu, jumlah kluster yang dipilih untuk analisis adalah tiga, yang memberikan keseimbangan antara penjelasan struktur data yang memadai dan kompleksitas model yang tidak berlebihan. Penurunan nilai inertia setelah $k=3$ tidak signifikan, sehingga jumlah kluster dipilih untuk menggambarkan struktur data tanpa menambah kompleksitas analisis[12].

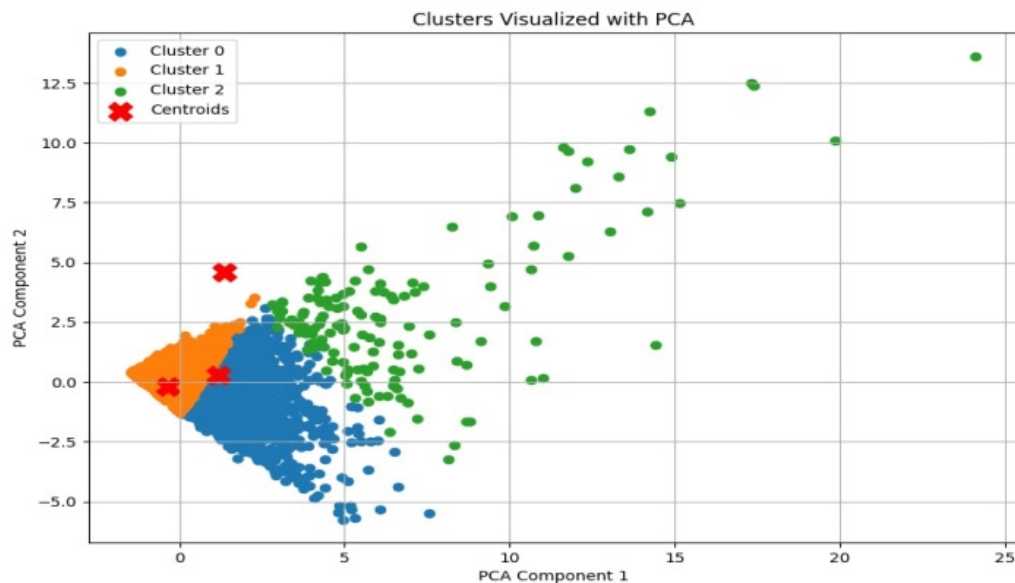
Setelah menentukan jumlah kluster, hasil clustering menunjukkan distribusi pelanggan ke dalam tiga kluster utama:

- 1) Kluster 0: Terdiri dari 1.968 pelanggan dengan saldo rekening sedang dan frekuensi pembelian yang tinggi, terutama didominasi oleh pembelian angsuran.
- 2) Kluster 1: Merupakan kluster terbesar dengan 6.833 pelanggan. Pelanggan dalam kluster ini memiliki saldo rendah hingga sedang, aktivitas pembelian yang merata, dan pembayaran yang cukup stabil.
- 3) Kluster 2: Kluster terkecil dengan 149 pelanggan. Pelanggan dalam kluster ini memiliki saldo rekening yang sangat tinggi dan batas kredit yang besar, tetapi frekuensi pembeliannya relatif rendah.

Tabel 2. Karakteristik Klaster Pelanggan

Atribut	Klaster 0	Klaster 1	Klaster 2	Interpretasi
Balance	Sedang	Rendah-Sedang	Sangat Tinggi	Klaster 2 memiliki saldo tertinggi
Purchases	Tinggi	Moderat	Rendah	Klaster 0 paling aktif berbelanja
Credit Limit	Moderat	Standar	Sangat Tinggi	Klaster 2 memiliki batas kredit terbesar
Payments	Tinggi	Moderat	Rendah	Sesuai dengan pola pembelian masing-masing klaster

Setelah menentukan jumlah klaster, hasil clustering menunjukkan distribusi pelanggan ke dalam tiga klaster utama. Untuk memahami perbedaan karakteristik antar klaster secara lebih mendalam, dilakukan analisis perbandingan atribut utama yang disajikan dalam Tabel 2 Analisis mendalam terhadap karakteristik yang ditampilkan dalam Tabel 2 memberikan wawasan strategis yang berharga. Pola yang menarik terlihat pada hubungan antara saldo dan aktivitas pembelian, di mana Klaster 0 dengan saldo sedang justru menunjukkan aktivitas pembelian tertinggi, mengindikasikan perilaku konsumtif yang aktif. Sebaliknya, Klaster 2 dengan saldo dan batas kredit tertinggi menunjukkan perilaku yang lebih konservatif dengan frekuensi pembelian yang rendah. Hal ini menunjukkan bahwa pelanggan dengan kapasitas finansial tinggi cenderung lebih selektif dalam melakukan transaksi, sementara pelanggan dengan saldo menengah lebih aktif dalam memanfaatkan fasilitas kredit yang tersedia.



Gambar 3. Visualisasi PCA

Visualisasi distribusi klaster menggunakan Principal Component Analysis (PCA) ditampilkan dalam Gambar 3. Visualisasi ini memperlihatkan bagaimana pelanggan tersebar dalam ruang dua dimensi berdasarkan komponen utama. Klaster 1 terlihat mendominasi, sementara klaster 2, meskipun kecil, menunjukkan karakteristik yang sangat spesifik. Centroid klaster, yang ditampilkan dalam bentuk simbol x, menunjukkan pusat dari masing-masing klaster. Distribusi ini membantu dalam mengidentifikasi segmen pelanggan yang berbeda berdasarkan pola keuangan mereka. Dari visualisasi ini, terlihat bahwa pelanggan dalam Klaster 2 memiliki karakteristik yang sangat berbeda dibandingkan Klaster 0 dan Klaster 1, meskipun jumlahnya lebih kecil.

3.3. Prediksi Pembayaran dengan Regresi Linier

Model prediktif regresi linier diterapkan untuk memprediksi pembayaran pelanggan (payments) menggunakan atribut balance, purchases, dan credit limit sebagai variabel independen. Hasil pelatihan model menghasilkan nilai Mean Squared Error (MSE) sebesar 4.796.998,52, yang menunjukkan bahwa terdapat kesalahan rata-rata yang signifikan antara prediksi dan nilai aktual. Nilai R-squared (R^2) sebesar 0,4276 menunjukkan bahwa model hanya mampu menjelaskan sekitar 42,76% variasi data pembayaran pelanggan.

Hasil ini mengindikasikan bahwa meskipun model memberikan wawasan tentang hubungan antara variabel utama dan pembayaran pelanggan, masih terdapat faktor lain yang memengaruhi pembayaran yang tidak tercakup dalam model. Penambahan variabel atau pendekatan model yang lebih kompleks, seperti regresi berganda atau algoritma ensemble, dapat meningkatkan akurasi prediksi[13].

3.4. Implikasi Bisnis

Analisis hasil clustering dan prediksi pembayaran dalam penelitian ini menunjukkan potensi besar dalam meningkatkan efektivitas strategi bisnis perusahaan, baik dalam hal pemasaran maupun pengelolaan risiko. Segmentasi pelanggan berdasarkan klaster memberikan pemahaman yang lebih mendalam tentang pola perilaku pelanggan dan kebutuhan spesifik masing-masing segmen. Klaster 0, yang didominasi oleh pelanggan dengan saldo rekening sedang dan frekuensi pembelian yang tinggi, dapat menjadi fokus utama dalam strategi retensi pelanggan. Kelompok ini memiliki potensi besar untuk tetap setia kepada perusahaan jika diberikan program loyalitas yang menarik, seperti diskon khusus, cashback, atau penawaran eksklusif. Aktivitas pembelian mereka yang konsisten menunjukkan bahwa mereka memiliki kepercayaan tinggi terhadap produk atau layanan perusahaan, sehingga mempertahankan pelanggan ini adalah langkah strategis yang dapat memperkuat pendapatan jangka panjang[14].

Sementara itu, Klaster 1 yang mencakup mayoritas pelanggan menunjukkan karakteristik saldo dan pembelian yang moderat. Dengan jumlah yang sangat besar, segmen ini menjadi kunci keberhasilan pemasaran massal perusahaan. Penawaran produk atau layanan dengan harga terjangkau, promosi paket bundling, atau bahkan penawaran berbasis kebutuhan spesifik mereka dapat meningkatkan daya tarik dan tingkat konversi di kelompok ini. Dalam konteks ini, analisis ini memberikan landasan yang kokoh bagi perusahaan untuk memaksimalkan potensi penjualan melalui pendekatan yang relevan dan terarah.

Berbeda dari dua klaster sebelumnya, Klaster 2 terdiri dari pelanggan yang memiliki saldo rekening dan batas kredit yang sangat tinggi, meskipun aktivitas pembelian mereka lebih rendah. Pelanggan dalam kelompok ini merupakan aset berharga bagi perusahaan karena mereka memiliki kapasitas finansial yang besar dan cenderung mencari layanan yang eksklusif. Oleh karena itu, perusahaan dapat mengembangkan strategi personalisasi yang lebih mendalam, seperti program premium, layanan konsultasi keuangan, atau akses ke produk-produk eksklusif. Dengan demikian, loyalitas pelanggan di segmen ini dapat ditingkatkan, sekaligus membuka peluang pendapatan tambahan dari transaksi bernilai tinggi[15].

Di sisi lain, hasil prediksi pembayaran juga memberikan wawasan strategis yang tak kalah penting. Dengan model yang mampu memprediksi potensi pembayaran pelanggan, perusahaan dapat mengidentifikasi pelanggan yang berisiko mengalami gagal bayar lebih awal. Informasi ini memungkinkan perusahaan untuk mengambil langkah mitigasi yang proaktif, seperti memberikan pengingat pembayaran, menawarkan opsi restrukturisasi kredit, atau memberikan insentif untuk pembayaran tepat waktu. Selain itu, hasil prediksi ini juga mendukung perusahaan dalam menetapkan kebijakan kredit yang lebih bijaksana, memastikan bahwa risiko kredit tetap terkendali tanpa mengorbankan pertumbuhan bisnis.

Secara keseluruhan, hasil analisis ini memberikan keunggulan kompetitif bagi perusahaan. Dengan memahami pola perilaku pelanggan dan memprediksi potensi pembayaran, perusahaan dapat mengoptimalkan strategi pemasaran, meningkatkan loyalitas pelanggan, dan memitigasi risiko keuangan. Dengan implementasi yang tepat, analisis ini tidak hanya memberikan manfaat jangka pendek tetapi juga mendukung keberlanjutan bisnis di masa depan. Penelitian ini

membuktikan bahwa pendekatan berbasis data dapat menjadi landasan yang kuat untuk pengambilan keputusan strategis yang lebih efektif dan berdampak[16].

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa penerapan teknik clustering dan model prediktif pada data pelanggan kartu kredit mampu mengungkap pola perilaku keuangan yang signifikan serta memberikan prediksi pembayaran yang bermanfaat untuk pengelolaan risiko dan strategi pemasaran. Melalui algoritma K-Means, pelanggan berhasil dikelompokkan ke dalam tiga klaster utama dengan karakteristik yang mencerminkan profil keuangan yang berbeda, seperti saldo rekening, frekuensi pembelian, dan batas kredit. Visualisasi menggunakan PCA mempermudah interpretasi hasil clustering, menunjukkan perbedaan jelas antara kelompok pelanggan.

Selain itu, model prediktif regresi linier yang digunakan untuk memperkirakan pembayaran pelanggan menunjukkan bahwa meskipun mampu memberikan gambaran pola hubungan antara variabel utama, tingkat akurasi model masih perlu ditingkatkan karena nilai MSE yang cukup tinggi dan R-squared yang hanya mampu menjelaskan sekitar 42,76% variasi data. Dengan demikian, penelitian ini berhasil menunjukkan potensi analisis data berbasis machine learning dalam mendukung pengambilan keputusan berbasis data, namun juga menyoroti perlunya pengembangan model yang lebih akurat untuk aplikasi yang lebih luas.

DAFTAR PUSTAKA

- [1] U. Srivastava and S. Gopalkrishnan, "Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks," *Procedia Comput. Sci.*, vol. 50, pp. 643–652, 2015, doi: 10.1016/j.procs.2015.04.098.
- [2] E. F. Qurani, "Dampak earning per share, debt to equity ratio, dan return on equity pada harga saham perusahaan: Studi empirik," *Implementasi Manaj. Kewirausahaan*, vol. 2, no. 1, pp. 38–49, Apr. 2022, doi: 10.38156/imka.v2i1.109.
- [3] M. B. Lendvai, I. Kovács, B. F. Balázs, and J. Beke, "Health and Environment Conscious Consumer Attitudes: Generation Z Segment Personas According to the LOHAS Model," *Soc. Sci.*, vol. 11, no. 7, p. 269, Jun. 2022, doi: 10.3390/socsci11070269.
- [4] M. Pradana, "Maximizing Strategy Improvement in Mall Customer Segmentation using K-means Clustering," *J. Appl. Data Sci.*, vol. 2, no. 1, Jan. 2021, doi: 10.47738/jads.v2i1.18.
- [5] D. Chandradhadinata and A. Sugiarto, "Analisis Produktivitas pada Produksi Dorokdok Mega Rasa dengan Metode American Productivity Center," *J. Kalibr.*, vol. 19, no. 1, pp. 7–15, Jan. 2022, doi: 10.33364/kalibrasi/v.19-1.1037.
- [6] C. B. Sadhewa and A. Z. Arifin, "Analisis Pengaruh Customer Satisfaction dan Customer Loyalty terhadap Financial Performance pada Layanan Bank Digital di Indonesia dari Sudut Pandang Konsumen," *J. Manaj. Bisnis Dan Kewirausahaan*, vol. 6, no. 6, pp. 615–620, Nov. 2022, doi: 10.24912/jmbk.v6i6.20687.
- [7] O. Aurelia, S. Hasbi, and F. Yetti, "Pengelolaan Pembiayaan Murabahah Nasabah Umkm Bprs Amanah Insani Di Masa Pandemi Covid-19: Management Of Murabahah Financing Umkm Customers Of Bprs Amanah Insani During The Covid-19 Pandemic," *NISBAH J. Perbank. SYARIAH*, vol. 7, no. 2, pp. 59–72, Dec. 2021, doi: 10.30997/jn.v7i2.4536.
- [8] A. Falakhi, "Pengolahan Data Pelanggan Dengan Teknik Clustering K-Means Di Aplikasi Weka," *J. Comput. Sci. Inf. Syst. J-Cosys*, vol. 3, no. 2, pp. 54–60, Jul. 2023, doi: 10.53514/jco.v3i2.394.
- [9] A. Hadi, "Segmentasi Pelanggan Internet Service Provider (ISP) Berbasis Pillar K-Means: segmentasi pelanggan, pilla K-Means, RFM," *J. Ilm. Teknol. Inf. Asia*, vol. 13, no. 2, pp. 151–159, Oct. 2019, doi: 10.32815/jitika.v13i2.413.
- [10] R. Liu, "Machine Learning Approaches to Predict Default of Credit Card Clients," *Mod. Econ.*, vol. 09, no. 11, pp. 1828–1838, 2018, doi: 10.4236/me.2018.91115.
- [11] I. Maryani, D. Riana, R. D. Astuti, A. Ishaq, Sutrisno, and E. A. Pratama, "Customer Segmentation based on RFM model and Clustering Techniques With K-Means Algorithm," in *2018 Third International Conference on Informatics and Computing (ICIC)*, Palembang, Indonesia: IEEE, Oct. 2018, pp. 1–6. doi: 10.1109/iac.2018.8780570.

-
- [12] E. Schubert, "Stop using the elbow criterion for k-means and how to choose the number of clusters instead," *ACM SIGKDD Explor. Newsl.*, vol. 25, no. 1, pp. 36–42, Jun. 2023, doi: 10.1145/3606274.3606278.
- [13] N. Shakhovska, N. Melnykova, V. Chopiyak, and M. Gregus MI, "An Ensemble Methods for Medical Insurance Costs Prediction Task," *Comput. Mater. Contin.*, vol. 70, no. 2, pp. 3969–3984, 2022, doi: 10.32604/cmc.2022.019882.
- [14] Y. S. Kartikawati, I. Istiatin, and S. Hartono, "Creating Customer Retention Throught Customer Incentives, Trust And Complaint Handling (At Bmt Amanah Ummah In Kartasura)," *Int. J. Econ. Bus. Account. Res. IJEBAR*, vol. 4, no. 02, Jun. 2020, doi: 10.29040/ijebar.v4i02.968.
- [15] Ridha Maisaroh and Maulida Nurhidayati, "Faktor-Faktor Yang Mempengaruhi Loyalitas Konsumen Toko Stars Madiun 2," *Niqosiya J. Econ. Bus. Res.*, vol. 1, no. 2, pp. 197–216, Dec. 2021, doi: 10.21154/niqosiya.v1i2.282.
- [16] R. Agusnawati, N. Nurfadillah, N. Wiradana, and A. Mukhtar, "Efektivitas Evaluasi Strategi dalam Manajemen Pengendalian Mutu Organisasi," *Indones. J. Innov. Multidisipliner Res.*, vol. 2, no. 1, pp. 87–105, Jun. 2024, doi: 10.69693/ijim.v2i1.148.