ANALISIS *CLUSTERING* SEGMENTASI PELANGGAN PADA CV PRIMA JAYA UTAMA MENGGUNAKAN METODE K-MEANS

Adelita Syaharani, Veri Ilhadi, Annisa Karima
Sistem Informasi, Universitas Malikussaleh
Jalan Batam, Blang Pulo, Muara satu – Lhokseumawe – Aceh (24352), Indonesia
adelita.210180138@mhs.unimal.ac.id

ABSTRAK

CV Prima Jaya Utama merupakan perusahaan yang bergerak dalam penjualan barang rumah tangga dan elektronik secara tunai maupun kredit. Perusahaan ini belum memiliki sistem segmentasi pelanggan yang terstruktur, sehingga strategi pemasaran dan kebijakan kredit belum berbasis analisis data. Penelitian ini bertujuan untuk mengelompokkan pelanggan berdasarkan karakteristik transaksi menggunakan metode K-Means Clustering. Data yang digunakan terdiri dari tiga variabel utama, yaitu harga angsuran, uang muka (DP), dan jumlah cicilan. Data dianalisis melalui tahapan preprocessing, meliputi pembersihan kolom, penanganan data kosong, dan normalisasi menggunakan StandardScaler. Proses clustering dilakukan dengan jumlah cluster sebanyak lima, berdasarkan hasil eksplorasi visual dan pertimbangan distribusi data. Evaluasi hasil clustering menggunakan Davies-Bouldin Index menghasilkan nilai sebesar 0,7878. Nilai ini termasuk kategori baik karena menunjukkan bahwa antar cluster memiliki pemisahan yang jelas dan distribusi data yang seragam dalam setiap cluster. Segmentasi ini memberikan gambaran yang representatif mengenai pola pelanggan dan dapat digunakan perusahaan untuk menyusun strategi pemasaran serta pengelolaan risiko kredit yang lebih terarah.

Kata kunci : *Segmentasi Pelanggan, K-Means, Clustering, Davies-Bouldin Index, CV Prima Jaya Utama*

1. PENDAHULUAN

Dalam era digital yang terus berkembang, bisnis dihadapkan pada perubahan yang cepat dan kompleks. Perkembangan teknologi yang pesat, perubahan perilaku konsumen, serta tantangan global yang semakin kompleks menuntut perusahaan untuk mengadopsi strategi adaptasi yang efektif [1]. Sebuah laporan McKinsey (2023) menunjukkan bahwa lebih dari 75% perusahaan global telah mulai menggunakan analisis data untuk mendukung pengambilan keputusan strategis.

CV Prima Jaya Utama merupakan salah satu perusahaan yang menerapkan strategi penjualan langsung melalui tim sales, dengan menyediakan berbagai produk konsumen seperti barang elektronik dan peralatan rumah tangga. Perusahaan ini menawarkan layanan pembelian secara tunai maupun kredit. Memberikan fleksibilitas kepada konsumen dalam memilih barang yang diminati, Salah satu cara untuk mencapai pemahaman yang lebih baik tentang pelanggan adalah melalui segmentasi pelanggan. Segmentasi pelanggan adalah proses membagi pasar menjadi kelompok-kelompok yang lebih kecil berdasarkan karakteristik tertentu, seperti demografi, perilaku, atau preferensi. Dengan melakukan segmentasi, perusahaan dapat mengembangkan strategi pemasaran yang lebih efektif dan efisien.

Dalam konteks ini, analisis segmentasi pelanggan sangat penting. Salah satu metode yang dapat digunakan untuk menganalisis segmentasi pelangan adalah K-Means clustering. Metode ini dipilih karena sifatnya yang mudah dipahami dan efektif dalam mengelompokkan data. K-Means merupakan algoritma yang umum digunakan dalam proses *clustering*. Algoritma ini mencari sejumlah klaster yang ditentukan berdasarkan kedekatan titik data satu sama lain [2]. Pengelompokan data menggunakan metode *clustering* dapat membantu CV Prima Jaya Utama Kisaran dalam mengidentifikasi kelompok pelanggan yang memiliki karakteristik serupa, sehingga dapat merumuskan strategi pemasaran yang lebih tepat sasaran. [3].

Berdasarkan latar belakang tersebut, penelitian ini akan berfokus pada "Analisis *Clustering* Segmentasi Pelanggan pada CV Prima Jaya Utama Kisaran Menggunakan Metode K-Means". Hasil penelitian diharapkan dapat memberikan rekomendasi strategis bagi manajemen dalam pengambilan keputusan bisnis yang lebih tepat dan efektif. Serta membantu perusahaan dalam mengembangkan segmentasi pelanggan yang lebih akurat.

2. TINJAUAN PUSTAKA

2.1. Segmentasi Pelanggan

Segmentasi pelanggan merupakan suatu proses dalam memecah pasar menjadi sejumlah kelompok yang lebih kecil dengan dasar karakteristik tertentu, seperti aspek demografis, perilaku, maupun preferensi konsumen [1]. Segmentasi pasar memungkinkan perusahaan untuk lebih memahami kebutuhan dan keinginan pelanggan, sehingga dapat merumuskan strategi pemasaran yang lebih efektif. Segmentasi yang tepat dapat membantu perusahaan dalam mengidentifikasi peluang pasar, meningkatkan kepuasan pelanggan, dan pada akhirnya meningkatkan profitabilitas.

2.2. Data Mining

Data mining adalah proses untuk menemukan pengetahuan baru dari sekumpulan data dengan

memanfaatkan algoritma dan teknik yang berasal dari disiplin ilmu seperti statistika, machine learning, dan sistem manajemen basis data [4]. Data mining digunakan untuk ekstraksi informasi penting yang tersembunyi dari dataset yang besar. Dengan adanya Data mining maka akan didapatkan suatu permata berupa pengetahuan di dalam kumpulan data yang banyak jumlahnya. Data mining merupakan proses untuk mengekstraksi atau menggali pengetahuan dari kumpulan data dalam jumlah besar. Istilah ini juga dikenal sebagai *Knowledge Discovery in Database* (KDD). Data mining sendiri merupakan salah satu tahap dalam proses KDD, yang mencakup beberapa langkah untuk menemukan pola atau informasi yang berguna dari data [5]:

a. *Data Selection*
   Menciptakan himpunan data target, pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (discovery) akan dilakukan. Hasil seleksi disimpan dalam suatu berkas, terpisah dari basis data operasional.
b. *Data Preprocessing*
   Pre-processing dan cleaning data merupakan operasi dasar yang dilakukan seperti penghapusan noise. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. Data bisa diperkaya dengan data atau informasi ekternal yang relevan.
c. *Data Transformation*
   Merupakan proses integrasi pada data yang telah dipilih, sehingga data sesuai untuk proses data mining. Merupakan proses yang sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.
d. *Data Mining*
   Pemilihan tugas data mining merupakan pemilihan goal dari proses KDD misalnya karakterisasi, klasifikasi, regresi, clustering, asosiasi, dan lain-lain. Pemilihan tugas data mining merupakan pemilihan goal dari proses KDD misalnya karakterisasi, klasifikasi, regresi, clustering, asosiasi, dan lain-lain. Pemilihan teknik, metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
e. *Interpretation/Evaluation*
   Yaitu penerjemahan pola-pola yang dihasilkan dari data mining. Pola informasi yang dihasilkan perlu ditampilkan dalam bentuk yang mudah dimengerti. Tahap ini melakukan pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

Tujuan dari data mining:

a. *Explonatory*, yaitu untuk menjelaskan beberapa kegiatan opservasi atau kondisi.
b. *Confirmatory*, yaitu untuk mengkonfirmasi suatu hipotesis yang telah ada.
c. *Exploratory*, yaitu untuk menganalisis data baru suatu relasi yang janggal.

2.3. Clustering

Clustering merupakan salah satu teknik dalam analisis data yang digunakan untuk mengelompokkan sekumpulan data ke dalam beberapa kelompok atau cluster berdasarkan kesamaan atribut atau karakteristik tertentu [2]. Tujuan dari *clustering* adalah untuk menemukan struktur alami dalam data, di mana data-data yang mirip ditempatkan dalam satu *cluster*, sedangkan data-data yang berlainan ditempatkan dalam *cluster* yang lain. Proses clustering mencakup pengelompokan data yang memiliki kedekatan dalam ruang atribut yang bersifat multidimensi. Hal ini dilakukan dengan menghitung tingkat kemiripan atau jarak antara data-data.

2.4. Algoritma K-Means

Algoritma K-Means adalah salah satu metode clustering yang populer dalam analisis data. Clustering merupakan metode untuk mengelompokkan data berdasarkan kemiripan atribut, di mana setiap objek dalam satu kelompok memiliki tingkat kesamaan yang lebih tinggi dibandingkan dengan objek-objek yang berada di kelompok lain [6]. Konsep dasar dari algoritma K-Means cukup sederhana, yakni dengan meminimalkan *Sum of Squared Error* (SSE) antara data-data objek melalui empat tahapan utama..

Adapun langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut:

a. Tentukan jumlah *cluster* (k)
b. Tentukan centroid awal secara random
c. Hitung jarak setiap data ke masing-masing centroid menggunakan rumus *Euclidean Distance*:

$$D(i,j) = \sqrt{(X_{1j} - X_{1j})^2 + (X_{2j} - X_{2j})^2 + \cdots + (X_{kj} - X_{kj})^2} \quad \ldots(1)$$

   D *(i,j)* = Jarak data ke i ke pusat cluster j
   Xki = Data ke i pada atribut data ke k
   Xkj = Titik pusat ke j pada atribut ke k
d. Kelompokkan data berdasarkan jarak terdekat ke centroid
e. Hitung ulang centroid baru berdasarkan rata-rata anggota *cluster*
f. Ulangi langkah 3-5 hingga centroid tidak berubah

2.5. Metode Elbow

Metode Elbow adalah salah satu teknik heuristik yang paling umum digunakan untuk menentukan jumlah *cluster* (k) yang optimal dalam algoritma *clustering* seperti K-Means [7] . Ide dasarnya adalah mencari nilai k di mana penambahan *cluster* lebih lanjut tidak lagi memberikan banyak informasi atau perbaikan yang signifikan dalam pengelompokan data.

Secara teknis, metode Elbow bekerja dengan menghitung *Sum of Squared Errors* (SSE), juga dikenal sebagai inertia, untuk berbagai nilai k. SSE

mengukur seberapa padat *(kohesif)* cluster yang terbentuk. Semakin kecil nilai SSE, semakin dekat titik-titik data ke centroid clusternya, yang menunjukkan *cluster* yang lebih padat [7].

Untuk penetuan jumlah klaster, menggunakan metode Elbow. Dengan mengukur jumlah kesalahan kuadrat atau *sum of squared error* (SSE) antara semua Objek dalam klaster Ci dan centroid Ci yang didefenisikan sebagai berikut [6]:

$$SSE = \sum_{K=1} \sum_{X_1 \in S_K} X_1 - C_K \quad \ldots(2)$$

Keterangan :
K = Jumlah klaster
Xi = Jumlah data
Ck = Jumlah clusteri pada cluster ke K

2.6. Evaluasi Kluster DBI

Indeks Davies-Bouldin adalah metrik yang digunakan untuk mengukur kualitas klastering dalam analisis data [6]. Tujuannya adalah untuk mengukur seberapa baik klaster yang dihasilkan oleh algoritma klastering memisahkan kelompok data yang berbeda dan mendekati pusat klasternya [8]. Semakin rendah nilai Davies-Bouldin Index, semakin baik klastering yang dihasilkan. Rumus Davies-Bouldin Index untuk menghitung kualitas klastering antara dua klaster Ci dan Cj adalah sebagai berikut:

$$R_{IJ} = \frac{S_I + S_J}{d_{IJ}} \quad \ldots(3)$$

Di mana:

a. Rij adalah nilai Davies-Bouldin Index antara klaster Ci dan Cj.
b. Si adalah ukuran dispersi atau sebaran data dalam klaster Ci, yang dapat dihitung dengan variasi atau jarak rata-rata antara titik data dalam klaster dengan pusat klaster mi.
c. Sj adalah ukuran dispersi atau sebaran data dalam klaster Cj, yang dapat dihitung dengan variasi atau jarak rata-rata antara titik data dalam klaster dengan pusat klaster mi.
d. dij adalah jarak antara pusat klaster mi dan mj

Langkah-langkah umum untuk menghitung Davies-Bouldin Index adalah sebagai berikut:

a. Hitung pusat klaster mi dan mj untuk masing-masing klaster Ci dan Cj.
b. Hitung ukuran dispersi Si untuk klaster Ci dan Sj untuk klaster Cj. Ini dapat dilakukan dengan menghitung jarak rata-rata antara titik-titik data dalam klaster dengan pusat klaster.
c. Hitung jarak dij antara pusat klaster mi dan mj.
d. Gunakan rumus Davies-Bouldin Index untuk menghitung nilai Rij antara klaster Ci dan Cj.
e. Ulangi langkah 1 hingga 4 untuk setiap pasangan klaster yang mungkin. Setelah semua nilai Rij dihitung, Davies-Bouldin Index keseluruhan dapat dihitung dengan rumus:

$$DB = \frac{1}{N} \sum_{i=1}^{N} max_{j \neq I} \left(R_{ij}\right) \quad \ldots(4)$$

Di mana:
DB adalah Davies-Bouldin Index keseluruhan.
N adalah jumlah total kluster

Tabel 1. Intepretasi nilai Davies-Bouldin Index

Davies-Bouldin Index	Intepretasi
<= 0,5	Cluster Sangat Baik
0,5 – 1,00	Cluster Baik
1,00 – 2,00	Cluster Cukup
≥ 2,00	Cluster Buruk

2.7. Penelitian Terdahulu

Ira Ariati dkk. (2023) dalam penelitiannya yang berjudul *"Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Pelanggan UHT Milk Greenfield"* menerapkan metode K-Means untuk mengelompokkan pelanggan produk susu UHT Greenfield. Hasil penelitian membentuk tiga kluster utama. Kluster Premium ditandai dengan pengiriman terbanyak ke wilayah Pamengkasan, di mana produk yang paling banyak dibeli adalah Greenfield UHT full cream 250 ml. Meskipun kuantitas pembeliannya tidak terlalu besar, kluster ini menghasilkan nilai penjualan yang signifikan. Kluster Menengah menunjukkan pengiriman dominan ke wilayah Jembrana, dengan produk yang banyak dibeli adalah Greenfield UHT full cream 125 ml, dan memiliki nilai penjualan terkecil karena pelanggan lebih fokus pada harga. Sementara itu, kluster Massal didominasi wilayah Jember, dengan pembelian berintensitas kecil tetapi kuantitas besar, yang turut memberikan kontribusi tinggi terhadap total nilai penjualan. Segmentasi ini menjadi dasar bagi perusahaan dalam menyusun strategi pemasaran yang lebih terarah dan sesuai karakteristik pelanggan di tiap segmen [9].

Ade Febriyanti dkk. (2024) dalam penelitiannya yang berjudul *"Analisis K-Means dalam Segmentasi Pasar Penggunaan Handphone di Lingkungan Mahasiswa STMIK Royal"* menggunakan algoritma K-Means untuk mengelompokkan 122 responden mahasiswa berdasarkan preferensi mereka terhadap smartphone. Penelitian ini menghasilkan tiga cluster utama: Cluster 1 (Fitur) terdiri dari 36 responden yang memprioritaskan harga, baterai, kamera, dan garansi; Cluster 2 (Produk) berisi 49 responden yang mengutamakan semua atribut kecuali garansi; dan Cluster 3 (Keunggulan) terdiri dari 37 responden yang fokus pada kamera, merek, dan RAM. Hasil ini menunjukkan pentingnya segmentasi pasar untuk pengembangan produk dan inovasi dalam industri smartphone [10].

Satria Ardi Perdana dkk. (2022) melakukan penelitian berjudul *"Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift"*. Penelitian ini memanfaatkan data transaksi pelanggan Alfagift pada bulan Juni 2021 dan membentuk tiga cluster berdasarkan usia dan metode pembayaran. Cluster 1 terdiri dari 7.219 pelanggan berusia 26–35 tahun yang dominan menggunakan metode pembayaran COD. Cluster 2 terdiri dari 6.902

pelanggan pada rentang usia yang sama tetapi lebih memilih pembayaran melalui Gopay. Cluster 3 berisi 5.371 pelanggan dengan karakteristik yang berbeda dari dua cluster sebelumnya. Penelitian ini memberikan kontribusi dalam memahami perilaku pelanggan untuk menyusun strategi pemasaran yang lebih efektif .

Nuril Huda Ahsina dkk. (2022) dalam penelitiannya yang berjudul *"Analisis Segmentasi Pelanggan Bank Berdasarkan Pengambilan Kredit dengan Menggunakan Metode K-Means Clustering"* menggunakan data sebanyak 1.000 nasabah. Hasil analisis menghasilkan empat cluster, yaitu: Cluster 1 dengan 286 nasabah (28,6%), Cluster 2 sebanyak 130 nasabah (13%), Cluster 3 sebanyak 542 nasabah (54,2%), dan Cluster 4 sebanyak 42 nasabah (4,2%). Pemilihan jumlah cluster didasarkan pada metode Elbow. Hasil segmentasi ini diharapkan dapat membantu bank dalam menyusun strategi pemasaran yang lebih tepat sasaran dalam penawaran kredit kepada nasabah [11].

Santi Ika Murpratiwi dkk. (2021) meneliti *"Analisis Pemilihan Cluster Optimal dalam Segmentasi Pelanggan Toko Retail"* pada UD. XYZ dengan menggunakan metode K-Means, K-Medoids, dan X-Means. Dari 153.492 data transaksi, dipilih 10.145 data pelanggan untuk proses segmentasi. Hasil analisis menunjukkan bahwa jumlah cluster terbaik adalah lima, dengan metode K-Medoids memiliki nilai Davies-Bouldin Index (DBI) rata-rata sebesar 0,540778, lebih baik dibandingkan metode lainnya. Kelima cluster tersebut dibagi berdasarkan tingkat loyalitas pelanggan: About To Sleep, Customer Needing Attention, Recent Customer, Potential Loyalist, dan Loyal Customers. Segmentasi ini digunakan untuk mendukung strategi pemasaran yang lebih efektif [12].

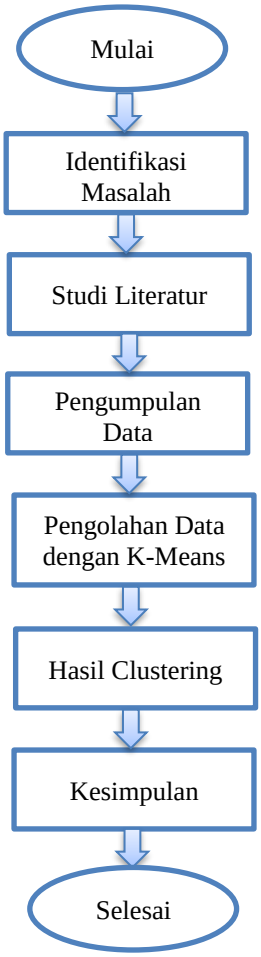
3. METODE PENELITIAN

Tahapan penelitian ini menjelaskan Langkah – Langkah yang ditempuh untuk mencapai tujuan penelitian ini. Penelitian ini merupakan jenis penelitian kuantitatif dengan pendekatan deskriptif. Pendekatan kuantitatif digunakan untuk mengukur dan menganalisis data dalam bentuk angka, yang memungkinkan peneliti untuk melakukan analisis statistik yang valid [5].

Adapun penjelasan dari setiap tahap adalah sebagai berikut:

3.1. Identifikasi Masalah

Tahapan awal dalam penelitian ini adalah mengidentifikasi permasalahan yang dihadapi oleh CV Prima Jaya Utama dalam mengelompokkan pelanggan berdasarkan karakteristik transaksi. Selama ini, perusahaan belum memiliki sistem yang mampu menganalisis pola pembelian pelanggan secara terstruktur, sehingga belum dapat melakukan segmentasi pelanggan secara efektif.



Gambar 1. Tahapan Penelitian

3.2. Studi Literatur

Peneliti mempelajari penelitian sebelumnya atau literatur yang relevan untuk memahami metode K-Means atau pendekatan lain yang telah digunakan untuk memecahkan masalah serupa. Tahap ini penting untuk mendapatkan dasar teori dan referensi metodologi.

3.3. Pengumpulan Data

Dalam penelitian ini, data dikumpulkan melalui beberapa metode:

a. Observasi Langsung pada CV Prima Jaya Utama
   Peneliti melakukan kunjungan langsung ke lokasi CV Prima Jaya Utama untuk mengamati proses operasional dan lingkungan kerja. Observasi ini bertujuan untuk memahami dinamika penjualan, interaksi antara karyawan dan pelanggan, serta bagaimana sistem informasi manajemen diterapkan di perusahaan.
b. Dokumentasi Data Penjualan Perusahaan
   Data penjualan yang diperlukan untuk analisis dikumpulkan melalui dokumentasi yang disediakan oleh perusahaan. Dokumentasi ini mencakup laporan pelanggan setiap tahun dengan mencakup nama pelanggan, tanggal tagihan, alamat, jenis produk (type), harga angsuran, DP, jumlah cicilan, serta status pembayaran.

c. Wawancara dengan Pihak Manajemen
Wawancara dilakukan dengan pihak manajemen CV Prima Jaya Utama untuk mendapatkan perspektif yang lebih dalam mengenai strategi penjualan, tujuan bisnis, dan tantangan yang dihadapi dalam mengelola data penjualan.

3.4. Pengolaan Data dengan K-Means

Data yang telah dikumpulkan kemudian diolah menggunakan metode K-Means *clustering* Tahap ini meliputi

a. Pembersihan data.
b. Menentukan jumlah *cluste*r (k)
c. Melakukan proses clustering untuk mengelompokkan data berdasarkan pola yang ditemukan.

3.5. Hasil Clustering

Hasil dari proses K-Means berupa cluster-cluster yang terbentuk, di mana setiap cluster menunjukkan kelompok data dengan karakteristik yang mirip. Setiap cluster ini memberikan wawasan penting mengenai pola pembelian, yang dapat digunakan untuk merencanakan strategi pemasaran secara lebih efektif.

Kesimpulan

Tahapan terakhir adalah menyusun kesimpulan berdasarkan hasil analisis dan pengujian data. Kesimpulan ini mencakup temuan-temuan utama seperti pola pembelian konsumen, segmentasi pelanggan berdasarkan kemiripan karakteristik transaksi, serta rekomendasi strategis bagi manajemen CV Prima Jaya Utama dalam merancang strategi pemasaran dan meningkatkan pelayanan kepada pelanggan.

4. HASIL DAN PEMBAHASAN

4.1. Penentuan Jumlah Cluster

Penentuan jumlah cluster optimal dilakukan dengan metode Elbow dan evaluasi Davies-Bouldin Index. Hasil analisis menunjukkan bahwa jumlah cluster optimal adalah lima, dengan nilai Davies-Bouldin Index sebesar 0,7878 yang menandakan kualitas cluster baik dan pemisahan antar cluster cukup jelas. Nilai ini menunjukkan bahwa cluster yang dihasilkan cukup baik dan dapat diterima. Semakin rendah nilai DBI, maka kualitas cluster yang terbentuk semakin baik karena menunjukkan bahwa jarak antar cluster cukup besar dan sebaran data dalam masing-masing cluster relatif kecil.

4.2. Karakteristik Cluster Berdasarkan Rata-rata Variabel

Tabel 2. Karakteristik Variabel

Variabel	Rendah	Sedang	Tinggi
Harga Angsuran	<159.370	159.370–194.785	>194.785
DP	<287.935	287.935–351.920	>351.920
Cicilan	<6.241	6.241–7.628	>7.628

untuk masing-masing variabel utama yang digunakan dalam analisis clustering, yaitu Harga Angsuran, DP, dan Cicilan. Batasan ini ditentukan berdasarkan rata-rata nilai seluruh data pelanggan, dengan toleransi ±10% sebagai rentang kategori sedang. Nilai di bawah 90% dari rata-rata diklasifikasikan sebagai rendah, nilai antara 90–110% sebagai sedang, dan nilai di atas 110% sebagai tinggi. Pendekatan ini digunakan agar penentuan karakteristik masing-masing cluster menjadi lebih objektif dan tidak hanya bergantung pada perbandingan antar cluster, tetapi berdasarkan distribusi nilai seluruh populasi data.

4.3. Interpretasi Cluster

Tabel 3. Interpretasi Cluster

Cluster	Harga Angsuran	DP	Cicilan
0	205.984	273.136	7.79
1	168.879	277.330	8.14
2	172.804	296.070	4.66
3	130.654	292.429	9.98
4	168.978	500.000	7.71

a. Cluster 0 – Pelanggan Stabil: Menunjukkan pelanggan dengan harga angsuran dan DP yang relatif tinggi serta jumlah cicilan menengah. Segmentasi ini menunjukkan pelanggan yang cenderung memilih pembayaran dengan DP dan angsuran sedang, sehingga memiliki risiko kredit yang relatif aman.
b. Cluster 1 – Pelanggan Reguler: Memiliki nilai rata-rata pada semua variabel di kisaran sedang. Cluster ini merupakan mayoritas pelanggan dengan pola pembayaran yang umum, tidak terlalu berisiko, dan paling mendominasi.
c. Cluster 2 – Pelanggan Cepat Lunas: Memiliki cicilan terendah dari semua cluster. DP dan harga angsuran juga tergolong sedang. Hal ini menunjukkan kelompok pelanggan yang memiliki kecenderungan melunasi kredit dengan cepat.
d. Cluster 3 – Pelanggan Berisiko: Terlihat dari cicilan yang sangat tinggi dan harga angsuran rendah, menunjukkan bahwa pelanggan ini cenderung memilih tenor panjang dan membayar angsuran kecil. Segmen ini berpotensi tinggi terhadap keterlambatan pembayaran.
e. Cluster 4 – Pelanggan Mapan: Memiliki DP tertinggi dibandingkan cluster lainnya dan cicilan sedang. Segmentasi ini menunjukkan pelanggan dengan daya beli tinggi yang mampu membayar uang muka besar dan melunasi kredit lebih cepat.

Hasil karakteristik ini disesuaikan dengan nilai rata-rata asli dari dataset serta dikaitkan dengan batas-batas kategori yang telah ditentukan berdasarkan jenis barang. Sehingga, kesimpulan tentang masing-masing cluster didasarkan pada perbandingan antara nilai hasil clustering dan klasifikasi tinggi, sedang, dan rendah dari setiap variabel. Dengan memahami karakteristik setiap cluster, perusahaan dapat menyesuaikan strategi

pemasaran, promosi, dan penilaian risiko kredit sesuai dengan tipe pelanggan masing-masing.

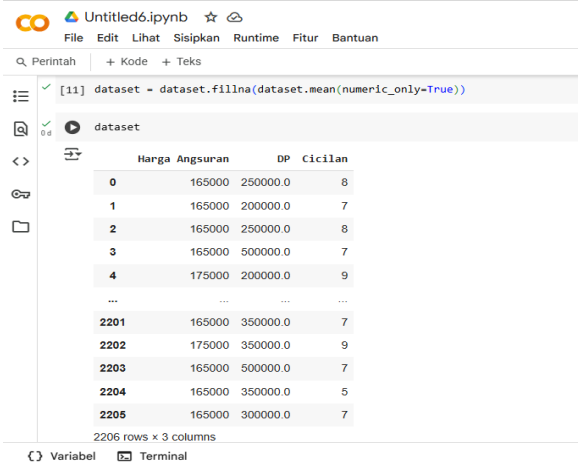
4.4. Menghapus kolom yang tidak digunakan

Gambar 2. Menghapus kolom yang tidak digunakan

Sebelum dilakukan proses klasterisasi, data terlebih dahulu dibersihkan dari atribut yang tidak relevan terhadap analisis clustering. Beberapa kolom dihapus karena tidak memberikan kontribusi langsung terhadap pemodelan segmentasi pelanggan.

Penghapusan kolom ini bertujuan untuk menyederhanakan data dan hanya menyisakan variabel-variabel numerik yang berkaitan langsung dengan analisis, seperti Harga Angsuran, DP, dan Cicilan, yang kemudian digunakan sebagai dasar dalam proses clustering menggunakan algoritma K-Means.

4.5. Mengisi Nilai yang Kosong dengan rata - rata



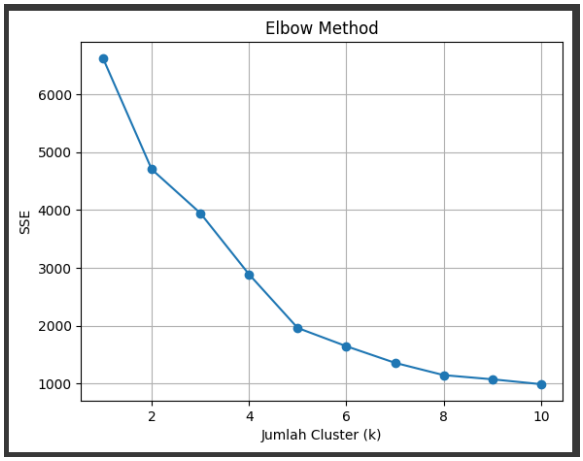
Gambar 3. Mengisi Nilai yang Kosong dengan rata – rata

Pengisian nilai kosong dilakukan dengan cara menghitung rata-rata dari setiap kolom numerik, kemudian mengganti nilai kosong pada kolom tersebut dengan nilai rata-rata tersebut.

Proses ini dilakukan secara otomatis oleh fungsi .fillna() pada pandas, dengan bantuan parameter dataset.mean(numeric_only=True) untuk memastikan hanya kolom numerik yang diproses

4.6. Menentukan nilai cluster k

Berdasarkan analisis grafik Elbow Method, terlihat bahwa penambahan jumlah cluster setelah k=5 tidak lagi memberikan penurunan SSE yang signifikan. Oleh karena itu, nilai k=5 dipilih sebagai jumlah cluster optimal untuk segmentasi pelanggan pada CV Prima Jaya Utama.



Gambar 4. Menentukan nilai cluster k

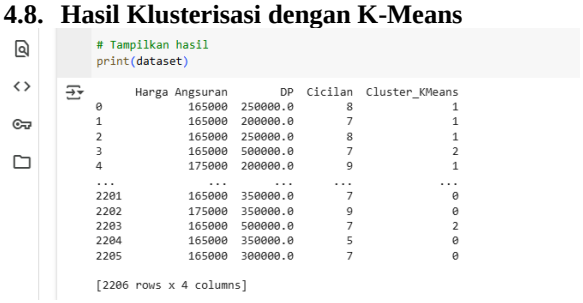
4.7. Klasterisasi Menggunakan K-Means

```
# 7. Proses K-Means
# kmeans = KMeans(n_clusters=optimal_k, random_state=42, n_init=10)
kmeans = KMeans(n_clusters=5, random_state=42)
y_predicted = kmeans.fit_predict(df_scaled)
```

Gambar 5. Klasterisasi Menggunakan K-Means

Langkah selanjutnya adalah menjalankan algoritma K-Means dengan jumlah cluster yang telah ditentukan sebanyak lima. Model K-Means kemudian dilatih pada data yang telah dinormalisasi untuk mengelompokkan pelanggan ke dalam lima segmen berdasarkan kemiripan karakteristik. Hasil dari proses ini adalah label cluster untuk setiap data pelanggan, yang digunakan untuk analisis lebih lanjut terhadap ciri khas masing-masing segmen.

4.8. Hasil Klusterisasi dengan K-Means



```
# Tampilkan hasil
print(dataset)
```

```
      Harga Angsuran       DP  Cicilan  Cluster_KMeans
0             165000  250000.0        8               1
1             165000  200000.0        7               1
2             165000  250000.0        8               1
3             165000  500000.0        7               2
4             175000  200000.0        9               1
...              ...       ...      ...             ...
2201          165000  350000.0        7               0
2202          175000  350000.0        9               0
2203          165000  500000.0        7               2
2204          165000  350000.0        5               0
2205          165000  300000.0        7               0

[2206 rows x 4 columns]
```

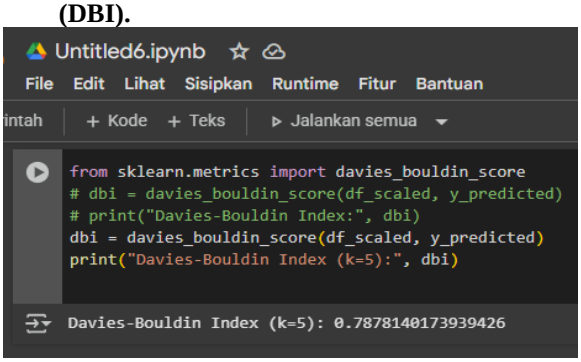
Gambar 6 Hasil Klusterisasi

Setelah proses clustering dilakukan dengan jumlah cluster sebanyak 5, hasil pengelompokan setiap data pelanggan ditampilkan dalam bentuk tabel. Tabel tersebut menunjukkan kolom Harga Angsuran, DP, Cicilan, serta Cluster_KMeans yang merupakan label cluster hasil klasifikasi dari algoritma K-Means.

Sebanyak 2.206 baris data telah berhasil diklasifikasikan ke dalam lima kelompok (cluster)

berbeda. Setiap baris merepresentasikan satu pelanggan dengan atribut nilai angsuran, jumlah uang muka, jumlah cicilan, dan informasi cluster tempat pelanggan tersebut termasuk. Nilai pada kolom Cluster_KMeans menunjukkan nomor cluster (misalnya 0, 1, 2, dst.) yang telah ditentukan berdasarkan kedekatan karakteristik antar data.

4.9. Evaluasi menggunakan Davies-Bouldin Index (DBI).

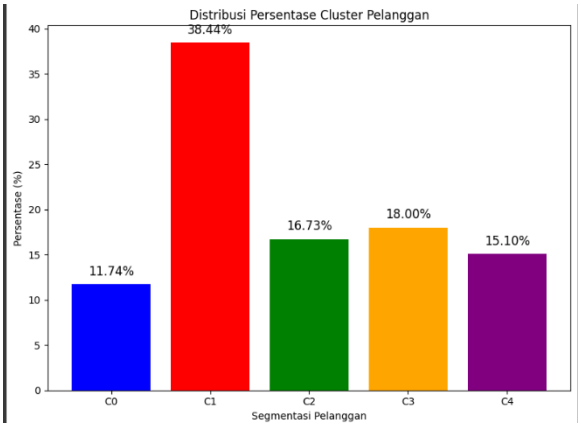


Gambar 7. Evaluasi menggunakan Davies-Bouldin Index (DBI)

Penjelasan pada gambar tersebut menunjukkan hasil evaluasi kualitas clustering menggunakan metrik Davies-Bouldin Index (DBI). DBI adalah salah satu metode evaluasi internal untuk menilai performa hasil clustering. Nilai DBI dihitung menggunakan fungsi davies_bouldin_score(df_scaled, y_predicted) yang mengukur seberapa baik pemisahan antar cluster dibandingkan dengan penyebaran data di dalam masing-masing cluster.

Pada hasil yang ditampilkan, nilai DBI = 0.7878140173939426 untuk jumlah cluster (k) = 5. Nilai ini berada pada rentang yang baik, karena semakin kecil nilai DBI, semakin baik hasil clustering. Artinya, masing-masing cluster yang terbentuk memiliki batas yang cukup jelas dan tidak saling tumpang tindih secara signifikan. Hasil ini juga memperkuat pemilihan jumlah cluster sebanyak 5 sebagai jumlah yang optimal dalam segmentasi pelanggan, karena menghasilkan pemisahan yang efisien dan representatif terhadap pola data.

4.10. Visualisasi Persentase Jumlah Per Cluster



Gambar 8. Visualisasi Persentase Jumlah per Cluster

Visualisasi Distribusi Persentase Cluster Pelanggan berdasarkan hasil klasterisasi K-Means yang dilakukan terhadap 2.206 data pelanggan CV Prima Jaya Utama. Grafik batang ini memperlihatkan proporsi jumlah pelanggan pada masing-masing cluster yang telah terbentuk.

Dari hasil visualisasi, dapat dilihat bahwa Cluster C1 mendominasi dengan persentase 38.44%, yang berarti sebagian besar pelanggan masuk ke dalam segmen ini. Sementara Cluster C0 merupakan kelompok dengan jumlah pelanggan paling sedikit, yaitu 11.74%. Cluster lainnya C2, C3, dan C4 memiliki persentase yang relatif seimbang, yaitu 16.73%, 18.00%, dan 15.10%.

Pentingnya grafik ini terletak pada kemampuannya dalam memberikan gambaran secara proporsional tentang persebaran pelanggan ke dalam masing-masing segmen yang terbentuk, sehingga dapat membantu manajemen dalam menentukan prioritas kebijakan atau strategi yang paling tepat bagi tiap kelompok pelanggan. Misalnya, cluster dengan jumlah besar mungkin membutuhkan pendekatan massal, sementara cluster kecil bisa difokuskan untuk layanan eksklusif atau strategi retensi tertentu.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan analisis yang telah dilakukan, penerapan metode K-Means clustering terbukti efektif dalam mengelompokkan pelanggan CV Prima Jaya Utama Kisaran berdasarkan variabel numerik, yaitu harga angsuran, uang muka (DP), dan jumlah cicilan. Proses klasterisasi dilakukan setelah melalui tahapan penting seperti pembersihan data, normalisasi, dan pengurangan dimensi agar hasil pengelompokan lebih akurat. Hasil klasterisasi membentuk lima segmen pelanggan dengan karakteristik berbeda: (1) Cluster 0 sebagai pelanggan stabil dengan harga angsuran tinggi dan DP seimbang, (2) Cluster 1 sebagai pelanggan reguler dengan perilaku pembayaran sedang dan konsisten, (3) Cluster 2 sebagai pelanggan cepat lunas dengan cicilan pendek dan pembayaran efisien, (4) Cluster 3 sebagai pelanggan berisiko yang memiliki cicilan panjang dan DP rendah, serta (5) Cluster 4 sebagai pelanggan mapan dengan DP tertinggi dan tenor cicilan yang pendek, menandakan daya beli yang kuat. Segmentasi ini dapat dijadikan dasar strategis bagi perusahaan dalam merancang kebijakan pemasaran, strategi penawaran kredit, serta pengelolaan risiko yang lebih terarah dan efektif.

Sebagai tindak lanjut, perusahaan disarankan untuk menggunakan hasil segmentasi ini sebagai acuan dalam mengambil keputusan pemasaran dan kebijakan kredit yang lebih tepat sasaran. Untuk meningkatkan akurasi klasterisasi, penelitian selanjutnya dapat mempertimbangkan penambahan variabel lain seperti riwayat keterlambatan pembayaran, lokasi pelanggan, atau jenis produk. Selain itu, integrasi hasil analisis segmentasi ke dalam sistem informasi internal perusahaan sangat

disarankan agar pengambilan keputusan dapat dilakukan secara otomatis dan berkelanjutan berdasarkan karakteristik tiap pelanggan.

DAFTAR PUSTAKA

[1] S. Rahmasari, "Strategi Adaptasi Bisnis di Era Digital: Menavigasi Perubahandan Meningkatkan Keberhasilan Organisasi," *Karimah Tauhid*, vol. 2, no. 3, pp. 622–636, 2023.

[2] Holwati, E. Widodo, and W. Hadikristanto, "Pengelompokan Untuk Penjualan Obat Dengan Menggunakan Algoritma K-Means," *Bull. Inf. Technol.*, vol. 4, no. 3, pp. 408–413, 2023, doi: 10.47065/bit.v4i3.848.

[3] S. Juniantoro and S. N. Yanti, "Sistem Informasi Inventory Stok Barang Pada Pt. Youngsun Berbasis Desktop," *Pros. Semin. SeNTIK*, vol. 7, no. 1, pp. 192–197, 2023, [Online]. Available: https://ejournal.jak-stik.ac.id/index.php/sentik/article/view/3426/688

[4] Hendry, "Data mining Prediksi Data Tingkat Malas Siswa Di Sekolah SMA," *Apl. dan Anal. Lit. Fasilkom UI*, vol. m, no. 1998, pp. 7–34, 2021, [Online]. Available: http://elib.unikom.ac.id/files/disk1/655/jbptunikompp-gdl-supriadini-32740-6-12.unik-i.pdf

[5] F. Agus, G. M. Putra, Z. A. Kamil, I. Arifin, and O. I. Gifari, "Peningkatan Kemampuan Analisis Statistik Kuantitatif Pada Riset Eksperimen Dengan Metode Workshop," *Plakat J. Pelayanan Kpd. Masy.*, vol. 4, no. 2, p. 243, 2022, doi: 10.30872/plakat.v4i2.8954.

[6] I. T. Umagapi, B. Umaternate, S. Komputer, P. Pasca Sarjana Universitas Handayani, B. Kepegawaian Daerah Kabupaten Pulau Morotai, and B. Riset dan Inovasi, "Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa," *Semin. Nas. Sist. Inf. dan Teknol.*, vol. 7, no. 1, pp. 303–308, 2023.

[7] R. Siagian, P. S. Pahala Sirait, and A. Halima, "E-Commerce Customer Segmentation Using K-Means Algorithm and Length, Recency, Frequency, Monetary Model," *J. INFORMATICS Telecommun. Eng.*, vol. 5, no. 1, pp. 21–30, Jul. 2021, doi: 10.31289/jite.v5i1.5182.

[8] Y. Sopyan, A. D. Lesmana, and C. Juliane, "Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1464–1470, 2022, doi: 10.47065/bits.v4i3.2697.

[9] S. Kasus, P. Uht, and M. Greenfield, "p-ISSN: 2774-6291 e-ISSN: 2774-6534 Available online at http://cerdika.publikasiindonesia.id/index.php/cerdika/index," vol. 3, no. 7, pp. 629–643, 2023.

[10] Ade Febriyanti, Putri Vina Bancin, and Siska Amanda, "Analisis K-Means dalam Segmentasi Pasar Penggunaan Handphone di Lingkungan Mahasiswa STMIK Royal," *J. Artif. Intell. Data Eng.*, vol. 1, no. 1, pp. 9–14, 2024.

[11] N. Ahsina, F. Fatimah, and F. Rachmawati, "Analisis Segmentasi Pelanggan Bank Berdasarkan Pengambilan Kredit Dengan Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. Infomasi Terap.*, vol. 8, no. 3, 2022, doi: 10.33197/jitter.vol8.iss3.2022.883.

[12] S. I. Murpratiwi, I. G. Agung Indrawan, and A. Aranta, "Analisis Pemilihan Cluster Optimal Dalam Segmentasi Pelanggan Toko Retail," *J. Pendidik. Teknol. dan Kejuru.*, vol. 18, no. 2, p. 152, 2021, doi: 10.23887/jptk-undiksha.v18i2.37426.