



K-Means-Based Customer Segmentation of Telkomsel Kartu Halo in the Sumbagsel

Segmentasi Pelanggan Kartu Halo Telkomsel Berbasis K-Means di Wilayah Sumbagsel

**Muhammad Bayu Samudra¹, Alifa Putri Shahabiyah², Aliya Faiza³, Edo Wicaksono⁴,
Ken Ditha Tania⁵, Allsella Meiriza⁶, Ahmad Rifai⁷**

^{1,2,3,4,5,6,7}Program Studi Sistem Informasi, Fakultas Ilmu Komputer,
Universitas Sriwijaya, Indonesia

E-Mail: ¹09031182328008@student.unsri.ac.id, ²09031282328116@student.unsri.ac.id,
³09031182328001@student.unsri.ac.id, ⁴09031282328067@student.unsri.ac.id, ⁵kenya.tania@gmail.com,
⁶allsella_meiriza@yahoo.co.id, ⁷ahmadrifai@ilkom.unsri.ac.id

Received Jan 27th 2026; Revised Feb 18th 2026; Accepted Mar 27th 2026; Available Online Apr 23th 2026

Corresponding Author: Aliya Faiza

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

As part of its daily operations, the telecommunications company handles a large amount of customer data, which increases as more people use digital services. However, this data is still often used for administrative purposes, so the potential to analyze it to understand customer behavior has not been fully realized. This research aims to group customers based on their service usage patterns in order to systematically identify the characteristics of each segment. The data used is customer data from postpaid Halo Card users in the Sumbagsel region for 2025, obtained from PT Telkomsel Smart Office Palembang, totaling 4,645 records. The approach used is Knowledge Discovery in Databases (KDD), which includes stages like preprocessing, transformation, and data mining. The clustering process is performed using the K-Means algorithm in RapidMiner. The research results show that customers can be divided into three groups with different service usage characteristics. The evaluation using the Davies-Bouldin Index indicates that the clustering model has good separation between groups. Scientifically, this study contributes to the application of clustering methods in customer segmentation within the telecommunications industry. In practice, the results can help create more targeted, data-driven marketing strategies.

Keywords: Customer Segmentation, K-Means Clustering, KDD, Regional Analysis, Telecommunications Data

Abstrak

Perusahaan telekomunikasi dalam menjalankan operasionalnya harus mengelola banyak data pelanggan, karena penggunaan layanan digital terus bertambah. Namun, data tersebut biasanya hanya digunakan untuk urusan administratif, sehingga kemampuan untuk menganalisis data guna memahami perilaku pelanggan belum sepenuhnya dimanfaatkan. Penelitian ini secara khusus bertujuan untuk mengelompokkan pelanggan berdasarkan pola penggunaan layanan guna mengidentifikasi karakteristik setiap segmen secara lebih sistematis. Data yang digunakan merupakan data pelanggan pascabayar Kartu Halo wilayah Sumbagsel tahun 2025 yang diperoleh dari PT Telkomsel Smart Office Palembang dengan total 4.645 data. Metode yang digunakan adalah *Knowledge Discovery in Databases* (KDD), yang terdiri dari beberapa tahap, yaitu *preprocessing*, *transformation*, dan *data mining*. Proses pengelompokan dilakukan dengan menggunakan algoritma *K-Means*, bantuan perangkat lunak RapidMiner. Penelitian menunjukkan bahwa pelanggan bisa dibagi menjadi tiga kelompok dengan kebiasaan menggunakan layanan yang berbeda. Evaluasi menggunakan *Davies-Bouldin Index* menunjukkan bahwa model *clustering* memiliki kualitas yang baik dalam memisahkan antarkelompok. Secara ilmiah, penelitian ini membantu dalam penerapan metode *clustering* untuk membagi pelanggan dalam industri telekomunikasi. Secara nyata, hasil ini bisa dipakai untuk membantu membuat strategi pemasaran yang lebih tepat dan didasarkan pada data.

Kata Kunci: Algoritma *K-Means*, Analisis Wilayah, KDD, Segmentasi Pelanggan, Telekomunikasi



1. PENDAHULUAN

Pelanggan memegang fungsi penting sebagai sumber keuntungan bagi perusahaan, terutama dalam bidang telekomunikasi yang sangat bergantung pada seberapa sering layanan digunakan oleh para pengguna. Kemajuan teknologi digital dan semakin meningkatnya permintaan masyarakat akan konektivitas jaringan telah mengubah cara orang menggunakan layanan telekomunikasi dengan cukup drastis. Saat ini, pelanggan tidak hanya melihat ketersediaan layanan komunikasi, tetapi juga mengharapkan kualitas jaringan, keandalan layanan, serta keberlangsungan akses yang mendukung berbagai kegiatan digital seperti komunikasi, hiburan, dan transaksi *online*. Situasi ini membuat persaingan antara penyedia layanan telekomunikasi di Indonesia menjadi semakin ketat, sehingga perusahaan perlu memahami kebutuhan pelanggan dengan lebih baik agar bisa menjaga loyalitas pengguna dan meningkatkan nilai bisnis secara berkelanjutan [1].

Besarnya volume data transaksi pelanggan yang dihasilkan oleh industri telekomunikasi mencerminkan meningkatnya intensitas interaksi masyarakat dengan layanan berbasis jaringan. Berdasarkan informasi dari Survei APJII: Gambaran Internet di Indonesia 2025, jumlah orang yang menggunakan internet di Indonesia mencapai 229,43 juta individu atau sekitar 80,66% dari jumlah keseluruhan penduduk nasional [2]. Tingginya penggunaan internet secara tidak langsung menyebabkan terakumulasinya data pelanggan dalam jumlah yang besar akibat aktivitas penggunaan layanan telekomunikasi. Data ini sebenarnya menyimpan berbagai informasi penting tentang perilaku pelanggan, cara mereka menggunakan layanan, dan pola transaksi yang bisa dijadikan acuan dalam menentukan kebijakan strategis perusahaan. Namun, dalam kenyataannya, banyak data operasional pelanggan yang masih sering digunakan hanya untuk tujuan administratif dan laporan rutin, sehingga potensi pengetahuan yang ada di dalamnya belum dimanfaatkan dengan optimal.

Pada layanan pascabayar seperti Kartu Halo PT Telkomsel, data dari transaksi pelanggan sangat berharga untuk memahami karakteristik dan tindakan pelanggan. Jika perusahaan tidak mengatur dan menganalisis data ini dengan tepat, perusahaan mungkin gagal menemukan tren penting dalam cara pelanggan menggunakan layanannya, yang dapat membantu dalam menciptakan strategi pemasaran yang lebih baik dan meningkatkan kualitas layanan. Akibatnya, penerapan teknik penambangan data sangat penting guna mengubah informasi pelanggan menjadi wawasan yang berharga. Metode yang populer digunakan dalam menganalisis segmen pelanggan adalah algoritma *K-Means Clustering*, yang merupakan teknik untuk mengkategorikan item ke dalam berbagai kelompok berdasarkan kesamaan fitur data mereka [3]. Dengan menggunakan teknik ini, pelanggan dapat diorganisir sesuai dengan kebiasaan pembelian mereka yang serupa, guna memperoleh pengelompokan yang lebih jelas berdasarkan cara mereka menggunakan layanan. Perusahaan dapat memanfaatkan wawasan ini untuk menciptakan rencana pemasaran yang lebih tepat sasaran dan meningkatkan praktik manajemen pelanggan mereka [4].

Beberapa riset terbaru menunjukkan bahwa algoritma *K-Means* sering dipakai dalam analisis segmentasi pelanggan di berbagai bidang industri. Penelitian yang dilakukan oleh Putri, Aldo, dan Ilham (2024) menggunakan *K-Means* yang digabungkan dengan pendekatan *Recency, Frequency, Monetary* (RFM) guna mengelompokkan pelanggan berdasarkan data transaksi ritel. Hasil penelitian menunjukkan bahwa pelanggan dapat dikelompokkan menjadi beberapa segmen seperti *VIP Customers, Loyal Customers, Potential Loyalists*, hingga *Dormant Customers*, sehingga perusahaan bisa menyiapkan rencana pemasaran yang lebih tepat untuk masing-masing kelompok pelanggan menurut karakteristik perilaku transaksi mereka. [5]. Penelitian lain oleh Yulisasih, Herman, dan Sunardi (2024) juga memanfaatkan *K-Means* mengelompokkan pelanggan pada sektor FMCG, menunjukkan bahwa metode ini mampu mengidentifikasi pola pembelian yang beragam. Hasilnya dapat dimanfaatkan untuk mendukung strategi promosi, pengelolaan stok, dan pengembangan produk yang lebih tepat sasaran [6].

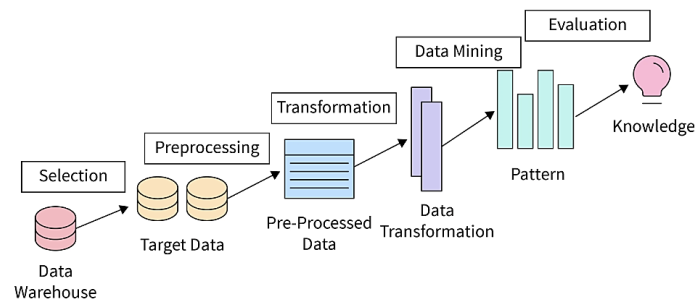
Selain itu, penelitian yang dilakukan oleh Vieri, Munandar, dan Srisulistiwati (2023) pada perusahaan telekomunikasi menunjukkan bahwa teknik *clustering* berbasis *machine learning* seperti *K-Means* dapat digunakan untuk menganalisis perilaku pelanggan berdasarkan data transaksi layanan. Hasil dari penelitian tersebut menunjukkan bahwa pemisahan pelanggan dapat memberikan informasi berharga tentang karakteristik penggunaan layanan dan memudahkan pengambilan keputusan bisnis suatu perusahaan yang lebih tepat, khususnya dalam merencanakan strategi pemasaran dan pengelolaan pelanggan yang didasarkan pada data [7]. Temuan tersebut sejalan dengan hasil kajian sistematis oleh Salminen *et al.* (2023) yang menerangkan algoritma *clustering*, terutama algoritma *K-Means*, adalah salah satu teknik paling umum dipakai dalam studi segmentasi pelanggan karena memiliki kemampuan yang sangat baik mengelompokkan data pelanggan dalam skala besar serta menghasilkan pola segmentasi yang mudah dipahami perusahaan [8].

Walaupun banyak studi sebelumnya telah membahas penggunaan *K-Means* untuk membagi pelanggan, sebagian besar penelitian tersebut masih berfokus pada sektor ritel, *e-commerce*, maupun layanan digital secara umum. Penelitian yang secara khusus menganalisis segmentasi pelanggan pada layanan telekomunikasi pascabayar dengan mempertimbangkan distribusi wilayah operasional cukup jarang dibahas. Untuk itulah, studi ini memiliki tujuan untuk menggunakan algoritma *K-Means Clustering* dalam membagi segmentasi pelanggan Kartu Halo PT Telkomsel di wilayah Sumbagsel berdasarkan data pelanggan tahun

2025. Penelitian ini mengelompokkan pelanggan ke dalam tiga kategori yaitu *High-tier Customers*, *Mid-tier Customers*, dan *Entry-tier Customers*, serta menganalisis distribusi pelanggan berdasarkan wilayah operasional. Pendekatan ini diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai karakteristik pelanggan baik dari sisi perilaku penggunaan layanan maupun persebaran geografisnya, sehingga hasil penelitian dapat dimanfaatkan sebagai dasar dalam mendukung pengambilan keputusan strategis berbasis data pada industri telekomunikasi.

2. METODOLOGI PENELITIAN

Penelitian ini mengaplikasikan pendekatan *Knowledge Discovery in Databases* (KDD) yang mencakup sejumlah tahapan utama dalam prosesnya, yaitu *selection* (pemilihan data), *preprocessing* (pra-pemrosesan data), *transformation* (transformasi data), *data mining*, dan *pattern evaluation* (evaluasi pola). Metode KDD banyak diterapkan dalam penelitian *data mining* karena mampu menggabungkan proses pengolahan data, mulai dari tahap pembersihan hingga evaluasi hasil analisis, secara sistematis dan terstruktur [9]. Rancangan alur penelitian yang digunakan dalam studi ini ditampilkan pada Gambar 1 sebagai gambaran tahapan yang dilakukan.



Gambar 1. Alur Proses Penelitian

Dalam penelitian ini, pendekatan *Knowledge Discovery in Databases* (KDD) digunakan bersamaan dengan metode *K-Means Clustering* untuk mengkategorikan segmen pelanggan dalam industri telekomunikasi. Metode *K-Means* dipilih karena dapat mengelompokkan data berdasarkan kemiripan dalam fitur-fitur dan telah umum diterapkan dalam penelitian sebelumnya [10] [11], serta dikenal sebagai metode *clustering* yang efisien dalam menangani data berukuran besar dan multidimensi [12]. Proses pengolahan data diawali dengan tahap pra-pemrosesan menggunakan *Microsoft Excel* yang meliputi identifikasi nilai kosong, pembersihan data tidak konsisten, serta normalisasi atribut agar data siap dianalisis lebih lanjut.

Setelah itu, metode pembentukan *cluster* dilakukan dengan *RapidMiner*, di mana jumlah *cluster* terbaik ditentukan dengan membandingkan beberapa skenario penentuan jumlah *cluster* dilakukan dengan menggunakan nilai *Davies-Bouldin Index* (DBI), yang berfungsi sebagai indikator kualitas hasil *clustering*. Hasil segmentasi kemudian divisualisasikan dan dianalisis untuk mengidentifikasi karakteristik masing-masing *cluster*. Informasi tersebut digunakan sebagai dasar untuk mendukung pengambilan keputusan yang berbasis data, terutama dalam menyusun rencana pemasaran yang lebih efektif bagi industri telekomunikasi.

2.1. Selection (Pemilihan Data)

Dalam tahap ini, fitur utama dipilih untuk mendukung proses pengelompokan pelanggan pada industri telekomunikasi. Data yang digunakan merupakan data sekunder pelanggan pascabayar Kartu Halo wilayah Sumbagsel tahun 2025 yang diperoleh dari PT Telkomsel Smart Office Palembang. Dataset awal disusun dalam bentuk data bulanan, kemudian digabungkan menjadi satu dataset terintegrasi dengan total 4.645 entri yang merepresentasikan pelanggan berdasarkan periode pengamatan. Distribusi *dataset* pelanggan berdasarkan periode tahun 2025 dapat dilihat pada Tabel 1.

Atribut penelitian yang digunakan mencakup: (1) periode yang diperoleh dari atribut *Tanggal Aktivasi* dan diubah ke dalam format bulan-tahun; (2) jenis perangkat yang berasal dari atribut *Jenis HP*; serta (3) indikator aktivitas layanan yang direpresentasikan melalui atribut yang tersedia pada dataset, seperti jenis paket yang digunakan pelanggan (*Paket*) dan lokasi aktivasi (*STORE*). Atribut aktivitas layanan dalam penelitian ini tidak diukur secara langsung melalui data penggunaan seperti konsumsi data atau durasi panggilan, melainkan menggunakan atribut transaksi sebagai pendekatan (*proxy*) untuk menggambarkan perilaku pelanggan. Atribut lain yang bersifat administratif dan tidak berkontribusi terhadap proses segmentasi tidak digunakan dalam analisis. Setiap atribut memiliki tipe data yang berbeda, dimana atribut periode bersifat temporal, sedangkan atribut jenis perangkat, paket, dan *store* merupakan atribut kategorikal. Untuk keperluan analisis *clustering*, atribut kategorikal tersebut kemudian ditransformasikan ke dalam

bentuk numerik menggunakan teknik *encoding* pada tahap *preprocessing*, sehingga data dapat diproses lebih lanjut menggunakan algoritma *K-Means*.

Tabel 1. Distribusi *Dataset* Pelanggan Berdasarkan Periode Tahun 2025

No	Periode (2025)	Jumlah Data
1	Januari	265
2	Februari	161
3	Maret	339
4	April	316
5	Mei	361
6	Juni	473
7	Juli	425
8	Agustus	427
9	September	465
10	Oktober	506
11	November	438
12	Desember	469
	Total	4.645

2.2. *Preprocessing* (Pra-pemrosesan Data)

Tahap *preprocessing* dijalankan agar data siap sebelum dilakukan proses *clustering*. Pembersihan awal dilakukan menggunakan *Microsoft Excel* yang meliputi pemeriksaan dan penanganan nilai kosong (*missing values*), penghapusan data duplikat, serta penyesuaian format dan konsistensi data antar periode. Setelah proses tersebut, *dataset* diimpor ke dalam *RapidMiner* untuk pengolahan lanjutan. Pada tahap ini dilakukan transformasi atribut kategorikal menjadi numerik menggunakan operator *Nominal to Numerical* serta normalisasi data dengan menerapkan metode *min-max normalization* agar setiap atribut memiliki skala yang seimbang. Tahapan ini bertujuan untuk memastikan bahwa data memiliki kualitas yang bersih dari noise dan memiliki kualitas yang baik, serta siap digunakan dalam proses *clustering* menggunakan algoritma *K-Means*, serta meningkatkan performa algoritma *clustering* dalam menghasilkan segmentasi yang lebih akurat [13].

2.3. *Transformation* (Transformasi)

Tahap transformasi dilakukan untuk menyesuaikan struktur data agar siap digunakan dalam proses *clustering*. Data hasil pra-pemrosesan diolah dengan mengubah atribut kategorikal menjadi numerik menggunakan operator *Nominal to Numerical* di *RapidMiner*. Selain itu, atribut yang tidak relevan terhadap proses segmentasi dihilangkan untuk meningkatkan fokus analisis. Penyesuaian skala data juga dilakukan melalui normalisasi agar setiap atribut memiliki kontribusi yang seimbang dalam proses pembentukan *cluster* menggunakan algoritma *K-Means*.

2.4. *Data mining*

Algoritma *K-Means* merupakan metode *clustering* berbasis partisi yang bekerja secara iteratif untuk mengklasifikasikan data ke dalam *cluster* tertentu berdasarkan jarak terdekat ke *centroid*. Tujuan utama dari algoritma ini adalah meminimalkan variasi dalam satu *cluster* (*intra-cluster*) serta memaksimalkan perbedaan antar *cluster* (*inter-cluster*). Secara matematis, *K-Means* meminimalkan fungsi objektif yang ditunjukkan pada Persamaan 1.

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (1)$$

di mana J merupakan total variasi dalam *cluster*, k adalah jumlah *cluster*, C_i adalah himpunan data pada *cluster* ke- i , dan μ_i adalah *centroid* dari *cluster* ke- i . Penelitian ini menggunakan *Euclidean Distance* sebagai metode pengukuran jarak yang dirumuskan pada Persamaan 2.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Persamaan tersebut digunakan untuk menghitung kedekatan antara data dengan *centroid*, sehingga proses pengelompokan dapat dilakukan secara optimal berdasarkan jarak minimum [12]. Proses *clustering* diawali dengan penentuan jumlah *cluster* (K), kemudian *centroid* awal diinisialisasi secara acak dengan *seed* tertentu untuk menjaga konsistensi hasil. Selanjutnya, dilakukan perhitungan jarak antara setiap data dengan *centroid* menggunakan *Euclidean Distance*, dan setiap data akan ditempatkan pada *cluster* dengan jarak terdekat. Setelah itu, *centroid* diperbarui berdasarkan nilai rata-rata dari seluruh data dalam masing-masing

cluster. Proses ini dilakukan secara iteratif hingga mencapai kondisi konvergen atau hingga mencapai batas maksimum iterasi yang telah ditentukan, yaitu sebanyak 100 iterasi. Dalam penelitian ini, jumlah *cluster* ditetapkan sebanyak $K = 3$ dengan mempertimbangkan kebutuhan segmentasi pelanggan menjadi tiga kelompok utama yang mudah diinterpretasikan. Hasil *clustering* kemudian dievaluasi menggunakan *Davies-Bouldin Index* (DBI), yang menunjukkan bahwa model yang dihasilkan memiliki kualitas *cluster* yang baik.

Atribut yang digunakan meliputi periode, jenis perangkat, serta indikator aktivitas layanan yang direpresentasikan melalui jenis paket dan lokasi aktivasi sebagai pendekatan terhadap perilaku pelanggan. Implementasi algoritma dilakukan menggunakan *RapidMiner* yang mencakup proses input data, transformasi atribut, penerapan *K-Means Clustering*, dengan pengukuran kualitas hasil menggunakan *Davies-Bouldin Index* (DBI). Hasil dari proses *clustering* menghasilkan beberapa segmen pelanggan dengan karakteristik yang berbeda, yang kemudian digunakan sebagai dasar dalam analisis segmentasi pelanggan untuk mendukung pengambilan keputusan berbasis data.

2.5. Pattern Evaluation (Evaluasi Pola)

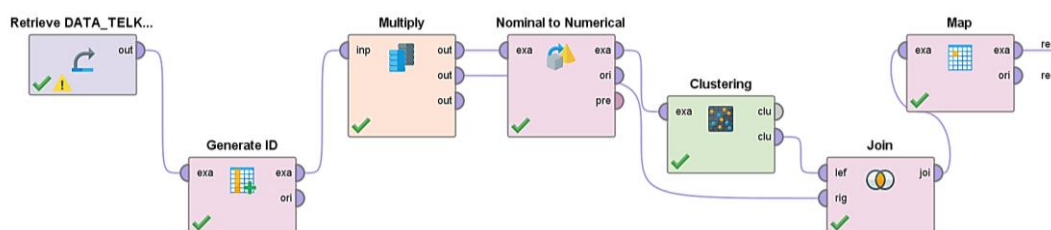
Evaluasi hasil *clustering* pada penelitian ini dilaksanakan dengan pendekatan *Davies-Bouldin Index* (DBI) untuk menilai seberapa baik segmentasi yang dihasilkan. Indeks DBI diterapkan untuk menilai tingkat pemisahan antarkluster serta keseragaman data di dalam masing-masing kluster, di mana semakin rendah nilai DBI, semakin baik kualitas kluster yang dihasilkan. Penggunaan DBI sebagai metode validasi telah banyak diterapkan dalam penelitian segmentasi pelanggan, baik untuk menentukan jumlah *cluster* optimal maupun untuk menilai kualitas hasil pengelompokan. Beberapa penelitian sebelumnya juga menunjukkan bahwa nilai DBI yang rendah berkorelasi dengan kualitas *cluster* yang lebih baik dan membantu dalam menentukan jumlah *cluster* yang paling sesuai [14].

3. HASIL DAN PEMBAHASAN

Bagian ini menggambarkan temuan yang diperoleh dari analisis dan wacana yang dihasilkan dari penerapan algoritma *K-Means* terhadap data pelanggan Kartu Halo Telkomsel tahun 2025. Proses *clustering* dilakukan untuk mengidentifikasi perbedaan pola penggunaan layanan antarpelanggan berdasarkan data transaksi yang dianalisis. Hasil pengelompokan menunjukkan bahwa pelanggan bisa dibagi menjadi tiga kelompok utama, yaitu *High-tier Customers*, *Mid-tier Customers*, dan *Entry-tier Customers*. Pembagian ini sejalan dengan pendekatan segmentasi nilai pelanggan yang umum digunakan dalam penelitian sebelumnya untuk membedakan tingkat kontribusi dan intensitas penggunaan layanan [15], [16]. Serta terbukti efektif dalam mengidentifikasi kelompok pelanggan berdasarkan perilaku dan nilai transaksi pelanggan [17]. Selain itu, distribusi *cluster* berdasarkan wilayah menunjukkan adanya perbedaan karakteristik pelanggan antardaerah. Temuan ini mengindikasikan bahwa faktor geografis turut memengaruhi pola penggunaan layanan, sebagaimana juga diungkapkan dalam penelitian terdahulu terkait segmentasi pelanggan berbasis wilayah [18].

3.1. Model Data Mining

Berdasarkan Gambar 2, urutan prosedur *data mining* yang dijalankan melalui penerapan *RapidMiner* digambarkan. Proses dimulai dengan operator *Retrieve* untuk memuat *dataset* hasil pra-pemrosesan ke dalam *workflow* analisis. Selanjutnya, operator *Generate ID* digunakan untuk menambahkan atribut identitas unik pada setiap data, kemudian dilakukan *Multiply* untuk menggandakan aliran data sesuai kebutuhan proses lanjutan.



Gambar 2. Model Data mining

Fase selanjutnya melibatkan transformasi *Nominal to Numerical*, yang memerlukan konversi atribut kategoris menjadi representasi numerik untuk memungkinkan pemrosesan dengan algoritma pengelompokan. Proses pengelompokan dijalankan menggunakan operator *Clustering* dalam hubungannya dengan algoritma *K-Means*, setelah itu *cluster* yang dihasilkan diintegrasikan kembali dengan data asli melalui aplikasi operator *Join*. Terakhir, operator *Map* digunakan untuk menyesuaikan label *cluster* menjadi kategori segmentasi pelanggan yang lebih interpretatif sehingga memudahkan analisis hasil.

3.2. Performance Vector

Berdasarkan Gambar 3, ditampilkan hasil evaluasi model *clustering* menggunakan *Performance Vector* pada *RapidMiner*. Nilai rata-rata jarak ke *centroid* (*Rata-rata. Within Centroid Distance*) sejumlah 867.802 menandakan bahwa variabilitas jarak data dari pusat *cluster* masih dikategorikan tinggi. Kondisi ini dapat disebabkan oleh distribusi data yang tidak merata atau adanya dominasi data pada *cluster* tertentu, sebagaimana dijelaskan dalam penelitian mengenai segmentasi pelanggan dengan algoritma *K-Means* [14].

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 867.802
Avg. within centroid distance_cluster_0: 868.176
Avg. within centroid distance_cluster_1: 0.000
Avg. within centroid distance_cluster_2: 0.000
Davies Bouldin: 0.295
```

Gambar 3. Performance Vector

Dalam algoritma *K-Means*, hasil pembagian data digambarkan melalui indeks kelompok, yaitu *Cluster_0*, *Cluster_1*, dan *Cluster_2*, guna menunjukkan cara data dibagi berdasarkan kedekatannya dengan pusat kelompok masing-masing. Penomoran tersebut tidak menunjukkan tingkatan tertentu, melainkan hanya sebagai identitas kelompok hasil iterasi algoritma. Setiap *cluster* memiliki karakteristik yang ditentukan oleh nilai rata-rata atribut anggotanya, sehingga interpretasi profil pelanggan dilakukan berdasarkan distribusi dan pusat data pada masing-masing *cluster*. Pendekatan ini biasa digunakan dalam studi segmentasi pelanggan yang menggunakan *K-Means* untuk mengenali perbedaan pola perilaku di antara kelompok data [14].

Rincian jarak rata-rata untuk setiap *cluster* adalah sebagai berikut:

1. *Cluster_0* memiliki jarak rata-rata paling tinggi, yaitu 868,176. Ini menunjukkan bahwa data di *cluster_0* lebih tersebar jika dibandingkan dengan *cluster-cluster* lain, sehingga dapat disimpulkan bahwa banyak data pelanggan terkumpul di kelompok utama ini.
2. *Cluster_1* dan *Cluster_2* memiliki rata-rata jarak 0,000, mengindikasikan bahwa jumlah data pada *cluster* ini sangat kecil atau memiliki kedekatan yang sangat tinggi terhadap *centroid*.

Davies-Bouldin Index (DBI) yang memiliki nilai sebesar 0,295 menunjukkan bahwa evaluasi pengelompokan termasuk dalam kategori baik. Besaran nilai tersebut menandakan bahwa kelompok yang terbentuk memiliki pemisahan yang cukup jelas dan kesamaan data yang cukup dalam setiap kelompok. DBI dihitung dengan membandingkan jarak dalam kelompok (variasi data di dalam satu kelompok) dan jarak antarkelompok (pemisahan antara kelompok-kelompok). Semakin kecil nilai DBI, semakin baik kualitas *cluster* yang dihasilkan [19]. Dalam studi ini, nilai DBI yang belum tinggi memperlihatkan bahwa hasil pemisahan dapat membedakan kelompok pelanggan dengan baik, meskipun masih ada perbedaan distribusi data di beberapa kelompok. Sejalan dengan temuan tersebut, penggunaan DBI sebagai matriks evaluasi dinilai terbukti mampu mengukur kualitas hasil *clustering* dan menjamin bahwa segmentasi pelanggan yang dihasilkan dapat digunakan sebagai dasar analisis profil pelanggan secara lebih terstruktur.

3.3. Visualisasi Chart

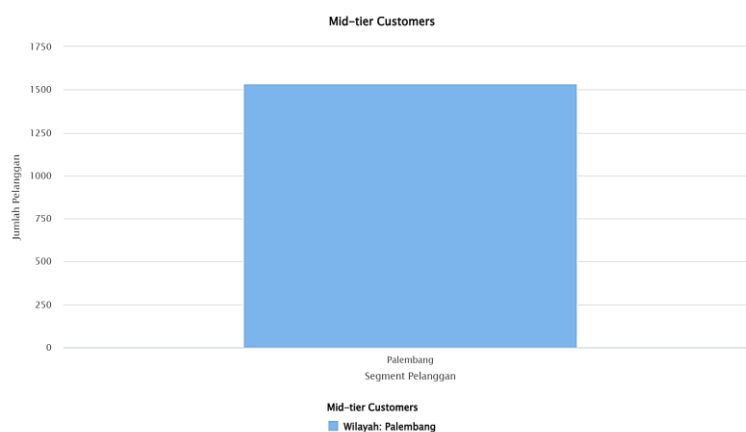
3.3.1. Visualisasi Entry-tier Customers



Gambar 4. Hasil Visualisasi Bar *Entry-tier Customers*

Visualisasi pada Gambar 4 menunjukkan distribusi pelanggan pada *cluster Entry-tier Customers* berdasarkan wilayah operasional di wilayah Sumbagsel. Grafik yang ditunjukkan menunjukkan bahwa jumlah pelanggan dalam kumpulan ini didominasi oleh kawasan Palembang, yang mempunyai jumlah pelanggan yang jauh lebih tinggi berbanding kawasan lain seperti Banyuasin, OKI, dan Prabumulih. Kondisi ini menunjukkan bahwa sebagian besar pelanggan dengan tingkat penggunaan layanan dasar berada di wilayah perkotaan yang memiliki tingkat aktivitas ekonomi dan kepadatan penduduk yang lebih tinggi. Sementara itu, wilayah lain memiliki jumlah pelanggan yang relatif lebih kecil, yang mengindikasikan adanya perbedaan karakteristik penggunaan layanan telekomunikasi di setiap wilayah. Temuan ini memperlihatkan bahwa faktor geografis turut mempengaruhi distribusi pelanggan pada segmen *Entry-tier Customers* dalam layanan Kartu Halo Telkomsel. Hal ini sejalan oleh penelitian yang menunjukkan bahwa kepadatan aktivitas ekonomi dan populasi berpengaruh terhadap intensitas penggunaan layanan digital [20].

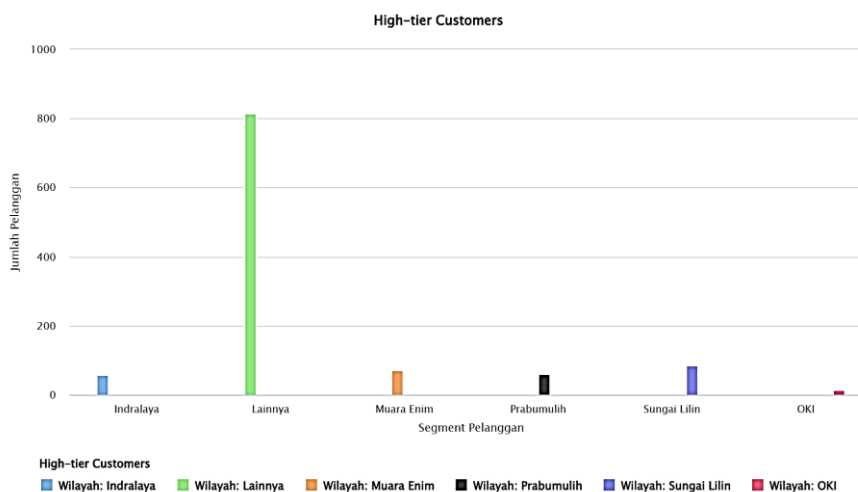
3.3.2. Visualisasi *Mid-tier Customers*



Gambar 5. Hasil Visualisasi Bar *Mid-tier Customers*

Visualisasi pada Gambar 5 menunjukkan distribusi pelanggan pada *cluster Mid-tier Customers* berdasarkan wilayah operasional di wilayah Sumbagsel. Berdasarkan hasil grafik, seluruh pelanggan pada segmen ini terkonsentrasi di wilayah Palembang, dengan jumlah yang cukup signifikan dibandingkan wilayah lainnya yang tidak menunjukkan adanya pelanggan dalam *cluster* tersebut. Hal ini menunjukkan bahwa pelanggan dengan tingkat penggunaan layanan menengah cenderung berada di wilayah perkotaan yang memiliki aktivitas ekonomi dan kebutuhan komunikasi yang lebih tinggi. Kondisi tersebut mengindikasikan bahwa wilayah Palembang menjadi pusat aktivitas pelanggan dengan pola penggunaan layanan yang lebih stabil dan moderat, sehingga berpotensi menjadi target pengembangan strategi pemasaran untuk meningkatkan penggunaan layanan menuju segmen pelanggan bernilai lebih tinggi.

3.3.3. Visualisasi *High-tier Customers*



Gambar 6. Hasil Visualisasi Bar *High-tier Customers*

Visualisasi pada Gambar 6 menunjukkan distribusi pelanggan pada *cluster High-tier Customers* berdasarkan wilayah operasional di wilayah Sumbagsel. Hasil grafik menunjukkan bahwa pelanggan pada segmen ini tersebar di beberapa wilayah, dengan jumlah terbesar berada pada kategori wilayah lainnya, diikuti oleh wilayah seperti Sungai Lilin, Muara Enim, Prabumulih, Indralaya, dan OKI dengan jumlah yang relatif lebih kecil. Pola distribusi ini menunjukkan bahwa pelanggan dengan tingkat penggunaan layanan yang tinggi tidak hanya terpusat pada wilayah perkotaan tertentu, tetapi juga tersebar di berbagai wilayah operasional lainnya. Hal ini mengindikasikan bahwa pelanggan bernilai tinggi memiliki karakteristik penggunaan layanan yang lebih luas secara geografis, sehingga perusahaan dapat mempertimbangkan strategi layanan premium dan peningkatan kualitas jaringan pada wilayah-wilayah yang menunjukkan potensi pelanggan dengan kontribusi penggunaan layanan yang tinggi.

3.3.4. Visualisasi Gabungan

Visualisasi ini menunjukkan distribusi jumlah pelanggan pada setiap *cluster* yang dihasilkan oleh algoritma *K-Means*. Perbedaan proporsi antarkluster menggambarkan adanya variasi karakteristik pelanggan dalam setiap kelompok. Berdasarkan Gambar 7, distribusi *cluster* pada masing-masing wilayah menunjukkan adanya variasi jumlah pelanggan pada setiap segmen. Wilayah dengan dominasi *cluster* tertentu mengindikasikan perbedaan karakteristik penggunaan layanan yang dipengaruhi oleh kondisi geografis dan aktivitas masyarakat setempat. Hasil ini sesuai dengan temuan studi-studi sebelumnya mengenai faktor-faktor regional, seperti tingkat urbanisasi dan aktivitas ekonomi, yang dapat memengaruhi pola penggunaan layanan telekomunikasi [18].



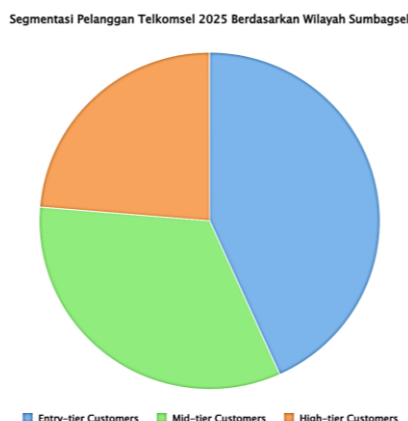
Gambar 7. Hasil Visualisasi Bar

Gambar 8 menyajikan distribusi proporsi pelanggan pada setiap *cluster* secara agregat. Hasil visualisasi menunjukkan bahwa *cluster Entry-tier Customers* mendominasi jumlah pelanggan secara keseluruhan, diikuti oleh *Mid-tier Customers* dan *High-tier Customers*. Dominasi pada segmen *Entry-tier Customers* menunjukkan bahwa sebagian besar pelanggan Kartu Halo berada pada tingkat penggunaan layanan yang relatif dasar dengan intensitas transaksi yang tidak terlalu tinggi. Kondisi ini mengindikasikan bahwa perusahaan memiliki basis pelanggan yang besar pada segmen penggunaan standar yang berpotensi untuk dikembangkan menjadi pelanggan dengan nilai penggunaan yang lebih tinggi. Sementara itu, segmen *Mid-tier Customers* menunjukkan kelompok pelanggan dengan tingkat penggunaan layanan yang lebih stabil dan moderat, yang dapat menjadi target strategi peningkatan pelayanan melalui program loyalitas atau penawaran paket layanan yang lebih optimal.

Di sisi lain, segmen *High-tier Customers* memiliki jumlah pelanggan yang lebih sedikit namun berpotensi memberikan kontribusi penggunaan layanan yang lebih tinggi. Oleh sebab itu, hasil pengelompokan ini dapat dimanfaatkan perusahaan guna merancang strategi pemasaran yang berbeda pada setiap segmen pelanggan, seperti peningkatan layanan premium bagi pelanggan bernilai tinggi serta strategi peningkatan penggunaan layanan bagi pelanggan pada segmen menengah dan dasar.

Dengan demikian, hasil visualisasi ini tidak hanya mendukung analisis segmentasi pelanggan berbasis data, tetapi juga memberikan gambaran yang lebih mendalam mengenai karakteristik masing-masing *cluster*. Pelanggan pada segmen *Entry-tier Customers* didominasi oleh pengguna dengan tingkat aktivitas layanan yang relatif rendah dan tersebar pada wilayah dengan kepadatan tertentu, sehingga berpotensi untuk ditingkatkan melalui strategi peningkatan penggunaan layanan. Sementara itu, *Mid-tier Customers*

menunjukkan pola penggunaan yang lebih stabil dan terkonsentrasi di wilayah perkotaan, sehingga memiliki peluang untuk dikembangkan menjadi pelanggan dengan nilai yang lebih tinggi. Di sisi lain, *High-tier Customers* meskipun jumlahnya lebih sedikit, memiliki kontribusi penggunaan layanan yang lebih besar dan tersebar di berbagai wilayah, sehingga perlu dipertahankan melalui layanan premium dan peningkatan kualitas jaringan.



Gambar 8. Hasil Visualisasi Pie

Hasil ini sesuai dengan studi sebelumnya yang menyebutkan bahwa segmentasi berbasis *clustering* mampu mengidentifikasi perbedaan perilaku pelanggan secara lebih spesifik dan mendukung pengambilan keputusan berbasis data [21] [22]. Oleh karena itu, hasil penelitian ini memiliki implikasi manajerial yang penting, di mana perusahaan dapat merancang strategi pemasaran yang berbeda pada setiap segmen, seperti program retensi untuk pelanggan bernilai tinggi, peningkatan loyalitas untuk pelanggan menengah, serta strategi promosi untuk mendorong peningkatan penggunaan pada pelanggan dengan aktivitas rendah.

4. KESIMPULAN

Penelitian ini berhasil menerapkan pendekatan *Knowledge Discovery in Databases* (KDD) yang dikombinasikan dengan algoritma *K-Means Clustering* untuk melakukan segmentasi pelanggan Kartu Halo Telkomsel berdasarkan data tahun 2025. Proses yang dilakukan meliputi tahap pemilihan data, pra-pemrosesan, hingga pemodelan dan evaluasi menggunakan *RapidMiner*. Hasil analisis menunjukkan bahwa pelanggan dapat dikelompokkan ke dalam tiga *cluster* utama, yaitu *Entry-tier Customers*, *Mid-tier Customers*, dan *High-tier Customers*, dengan nilai *Davies-Bouldin Index* (DBI) sebesar 0,295 yang mengindikasikan kualitas *clustering* yang baik. Segmentasi yang dihasilkan mampu menggambarkan perbedaan karakteristik pelanggan secara jelas, dimana sebagian besar pelanggan berada pada kategori *Entry-tier Customers*, diikuti oleh *Mid-tier Customers*, sementara *High-tier Customers* memiliki jumlah lebih sedikit namun berpotensi memberikan kontribusi nilai yang lebih tinggi bagi perusahaan.

Dengan demikian, tujuan penelitian untuk mengidentifikasi segmentasi pelanggan berbasis pola layanan menggunakan metode *K-Means* telah tercapai. Secara praktis, hasil penelitian ini dapat dimanfaatkan oleh perusahaan, khususnya PT Telkomsel, sebagai dasar dalam merancang strategi pemasaran yang lebih efektif dan terarah. Perusahaan dapat mengoptimalkan personalisasi layanan, memprioritaskan pengelolaan pelanggan bernilai tinggi, serta menyusun pendekatan pemasaran yang berbeda pada setiap segmen. Selain itu, informasi distribusi *cluster* juga dapat digunakan untuk mendukung pengambilan keputusan terkait pengembangan layanan di wilayah yang memiliki potensi pertumbuhan pelanggan. Untuk penelitian selanjutnya, disarankan untuk mengembangkan model segmentasi dengan menambahkan atribut data yang lebih beragam agar hasil analisis menjadi lebih komprehensif. Selain itu, penggunaan dan perbandingan dengan metode *clustering* lain seperti *Hierarchical Clustering*, DBSCAN, atau *Fuzzy C-Means* perlu dilakukan untuk memperoleh hasil segmentasi yang lebih optimal dan akurat.

REFERENSI

- [1] R. I. Sujono *et al.*, "Maintaining sustainable use of the Indonesian telecommunications provider," *Jurnal Studi Komunikasi (Indonesian Journal of Communications Studies)*, vol. 8, no. 1, pp. 42–52, Mar. 2024, doi: 10.25139/jsk.v8i1.6246.
- [2] Asosiasi Penyelenggara Jasa Internet Indonesia, "Survei Internet APJII 2025," 2025. [Online]. Available: <https://survei.apjii.or.id/> (accessed Feb. 15, 2026).
- [3] A. P. Pramudiansyah, "Segmentasi pelanggan menggunakan algoritma *K-Means* berdasarkan model Recency Frequency Monetary," *Jurnal Ilmiah Ilmu Komputer*, vol. 7, no. 2, pp. 6–19, Sep. 2021, doi:

- 10.35329/jiik.v7i2.201.
- [4] P. M. K. Mado and H. Hendry, "Implementasi algoritma *clustering K-Means* untuk segmentasi pelanggan di e-commerce," *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, vol. 6, no. 3, pp. 1680–1686, Sep. 2025, doi: 10.63447/jimik.v6i3.1563.
- [5] Y. Putri, D. Aldo, and W. Ilham, "Retail marketing strategy optimization: Customer segmentation with artificial intelligence integration and *K-Means clustering*," *Sinkron*, vol. 8, no. 4, pp. 2155–2163, Oct. 2024, doi: 10.33395/sinkron.v8i4.14000.
- [6] B. N. Yulisasih, H. Herman, and S. Sunardi, "*K-Means clustering* method for customer segmentation based on potential purchases," *Jurnal ELTIKOM*, vol. 8, no. 1, pp. 83–90, Jun. 2024, doi: 10.31961/eltikom.v8i1.1137.
- [7] J. K. Vieri, T. A. Munandar, and D. B. Srisulistiwati, "Exclusive *clustering* technique for customer segmentation in national telecommunications companies," *International Journal of Information Technology and Computer Science Applications*, vol. 1, no. 1, pp. 51–57, Jan. 2023, doi: 10.58776/ijitcsa.v1i1.19.
- [8] J. Salminen, M. Mustak, M. Sufyan, and B. J. Jansen, "How can algorithms help in segmenting users and customers? A systematic review and research agenda for algorithmic customer segmentation," *Journal of Marketing Analytics*, vol. 11, no. 4, pp. 677–692, Jul. 2023, doi: 10.1057/s41270-023-00235-5.
- [9] D. Setiono, "Penerapan Model Choo Sense Making Untuk Meningkatkan Sharing Knowledge Pada Laboratorium Media Komunikasi," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 4, pp. 1079–1085, Oct. 2020, doi: 10.30865/mib.v4i4.2360.
- [10] F. Nazihah, A. Danniswara, and A. Wibowo, "Analisis segmentasi pelanggan dengan algoritma *K-Means* pada data penjualan," *Jurnal Algoritma*, vol. 22, no. 2, Nov. 2025, doi: 10.33364/algoritma/v.22-2.2489.
- [11] Refri Martiansah, Siti Monalisa, Fitriani Muttakin, and Mona Fronita, "Customer Segmentation Analysis Through RFM-D Model and *K-Means* Algorithm," *Jurnal Sistem Cerdas*, vol. 8, no. 1, pp. 1–11, Apr. 2025, doi: 10.37396/jsc.v8i1.504.
- [12] J. Han, M. Kamber, and J. Pei, *Data mining: Concepts and Techniques*. Elsevier, 2011.
- [13] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data mining*. Springer, 2014.
- [14] R. Ramdani, N. Suarna, I. Ali, and D. I. Efendi, "Penerapan algoritma *K-Means* dalam analisis data kependudukan untuk optimalisasi pengelompokan di Desa Pasawahan," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5639.
- [15] Y. Syahra, A. Fadlil, and H. Yuliansyah, "Customer segmentation using RFM and *K-Means clustering* to support CRM in retail industry," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 9, no. 3, pp. 1120–1131, 2025.
- [16] A. A. Rahma, A. Faqih, and A. R. Rinaldi, "Optimalisasi strategi pemasaran melalui segmentasi pelanggan dengan analisis RFM dan algoritma *K-Means* untuk bisnis ritel," *JIKO (Jurnal Informatika dan Komputer)*, vol. 9, no. 2, p. 338, Jun. 2025, doi: 10.26798/jiko.v9i2.1737.
- [17] A. Sakinah and D. S. Awaliyah, "Optimization e-commerce consumer segmentation based on *K-Means clustering* and machine learning," *Journal of Mathematics, Computations and Statistics*, vol. 8, no. 2, pp. 606–619, Sep. 2025, doi: 10.35580/jmathcos.v8i2.9548.
- [18] R. Perdhana and J. Heikal, "Enhancing customer segmentation in online transportation services using *K-Means clustering* and RFM model," *Indonesian Interdisciplinary Journal of Sharia Economics*, vol. 7, no. 2, pp. 1012–1024, 2024, doi: 10.31538/ijjse.v7i2.4851.
- [19] A. Julian, "Segmentasi pelanggan menggunakan algoritma *K-Means* pada jaringan telekomunikasi untuk optimalisasi strategi pemasaran," *Jurnal Teknologi Terpadu*, vol. 13, no. 1, pp. 94–100, Apr. 2025, doi: 10.32487/jtt.v13i1.2436.
- [20] "Customer segmentation using *K-Means clustering* for targeted marketing in banking," *Zenodo*, 2024, doi: 10.5281/zenodo.13627403.
- [21] S. F. Djun, I. G. A. Gunadi, and S. Sariyasa, "Analisis segmentasi pelanggan pada bisnis dengan menggunakan metode *K-Means clustering* pada model data RFM," *JTIM: Jurnal Teknologi Informasi dan Multimedia*, vol. 5, no. 4, pp. 354–364, Feb. 2024, doi: 10.35746/jtim.v5i4.434.
- [22] D. A. Awaliyah, B. Prasetyo, R. Muzayanah, and A. D. Lestari, "Optimizing customer segmentation in online retail transactions through the implementation of the *K-Means clustering* algorithm," *Scientific Journal of Informatics*, vol. 11, no. 2, pp. 539–548, Jun. 2024, doi: 10.15294/sji.v11i2.6137.