



Application of the Recency, Frequency, Monetary Model for Segmentation of Indibiz Customer Behavior Patterns

Penerapan Model *Recency, Frequency, Monetary* untuk Segmentasi Pola Perilaku Pelanggan *Indibiz*

**Aulia Pinkasari^{1*}, Meiyyin Monica Amilia Putri², Gibral Abdurahman³,
Ahmad Fadhil Rizqi⁴, Ken Ditha Tania⁵, Allsela Meiriza⁶, Ahmad Rifai⁷**

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

E-Mail: ¹09031182328011@student.unsri.ac.id, ²09031182328097@student.unsri.ac.id,
³09031282328075@student.unsri.ac.id, ⁴09031282328058@student.unsri.ac.id,
⁵kenya.tania@gmail.com, ⁶allsela@unsri.ac.id, ⁷ahmadrifai@ilkom.unsri.ac.id

Received Feb 01st 2026; Revised Feb 26th 2026; Accepted Mar 27th 2026; Available Online Apr 23th 2026
Corresponding Author: Aulia Pinkasari
Copyright ©2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

PT Telkom Indonesia Witel Sumbagsel is facing the "Data Rich, Information Poor" (DRIP) condition, characterized by an abundance of transaction data that has not been optimally utilized as a basis for decision-making in the Payment Collection unit. This study applies the Knowledge Discovery in Databases (KDD) methodology using the RFM (Recency, Frequency, Monetary) model and the K-Means Clustering algorithm to identify patterns in customer payment behavior. The dataset consists of 46,355 transactions from January to December 2025. The optimal number of clusters was determined using the Elbow method, resulting in four customer segments ($k=4$). The Silhouette Coefficient evaluation yielded a score of 0.3463, indicating acceptable cluster quality. The segmentation grouped customers into loyal, potential, standard, and churn-risk categories, thereby supporting the development of more targeted, data-driven billing strategies. This study demonstrates that integrating RFM and K-Means effectively transforms transaction data into actionable insights for telecommunications billing management.

Keywords: Customer Segmentation, Data Mining, Indibiz, K-Means Clustering, RFM

Abstrak

PT Telkom Indonesia Witel Sumbagsel menghadapi kondisi *Data Rich, Information Poor* (DRIP), yaitu melimpahnya data transaksi yang belum dimanfaatkan secara optimal sebagai dasar pengambilan keputusan pada unit *Payment Collection*. Penelitian ini menerapkan metodologi *Knowledge Discovery in Databases* (KDD) menggunakan model RFM (*Recency, Frequency, Monetary*) dan algoritma *K-Means Clustering* untuk mengidentifikasi pola perilaku pembayaran pelanggan. *Dataset* terdiri dari 46.355 transaksi periode Januari–Desember 2025. Jumlah *cluster* optimal ditentukan menggunakan metode *Elbow* dan menghasilkan empat segmen pelanggan ($k=4$). Evaluasi menggunakan *Silhouette Coefficient* memperoleh nilai 0,3463 yang menunjukkan kualitas kluster yang dapat diterima. Hasil segmentasi mengelompokkan pelanggan ke dalam kategori *loyal*, *potential*, *standard*, dan *churn-risk*, sehingga mendukung penyusunan strategi penagihan yang lebih tepat sasaran dan berbasis data. Penelitian ini menunjukkan bahwa integrasi RFM dan K-Means efektif dalam mengubah data transaksi menjadi wawasan praktis bagi manajemen penagihan telekomunikasi.

Kata Kunci: *Data Mining*, *Indibiz*, *K-Means Clustering*, Segmentasi Pelanggan, RFM

1. PENDAHULUAN

Pertumbuhan volume data yang tinggi dalam organisasi tidak selalu berbanding lurus dengan peningkatan kualitas pengambilan keputusan. Banyak organisasi menghadapi fenomena *Data Rich, Information Poor* (DRIP), di mana mereka memiliki "gudang data" yang terus meningkat namun sulit mendapatkan informasi yang bermanfaat untuk pengambilan keputusan [1]. Fenomena ini semakin relevan pada perusahaan penyedia layanan digital berskala besar, di mana volume data pelanggan terus meningkat



seiring dengan kompleksitas layanan yang ditawarkan. PT Telkom Indonesia dikenal sebagai salah satu perusahaan terbesar di Indonesia yang bergerak di bidang penyedia layanan internet dan memiliki berbagai cabang di berbagai regional (wilayah), salah satunya Telkom Witel Sumbagsel. PT Telkom terus berinovasi dalam mengembangkan produknya, salah satunya adalah layanan *Indibiz*.

Indibiz merupakan layanan internet dan telepon sebagai upaya untuk memfasilitasi digitalisasi bisnis mikro kecil menengah (UMKM). Salah satu unit yang bertanggung jawab adalah unit *Payment Collection*. Masalah praktis yang dihadapi unit *Payment Collection* di lingkungan telekomunikasi adalah ketidakmampuan menginterpretasikan volume data yang besar menjadi *knowledge actionable* yang dapat langsung dimanfaatkan dalam meningkatkan kinerja dan produktivitas operasional, seperti tidak adanya informasi detail mengenai karakteristik pelanggan. Merespons tantangan tersebut, penelitian kami mengusulkan untuk melakukan analisis karakteristik pelanggan sebagai basis segmentasi pelanggan [2]. Tujuan dari hal tersebut adalah untuk mengetahui karakteristik pelanggan dan memberikan aksi solusi strategis yang tepat untuk diterapkan untuk keuntungan perusahaan.

Studi terbaru menunjukkan bahwa pendekatan *customer segmentation* berbasis *Knowledge Discovery in Databases* (KDD) mampu menemukan kelompok pelanggan dengan karakter perilaku yang sama berdasarkan *recency*, *frequency*, dan *monetary value* transaksi mereka, sehingga menghasilkan keputusan strategis yang lebih baik dalam pemanfaatan perilaku pelanggan untuk menentukan strategi penagihan [3]. Pendekatan ini memungkinkan organisasi untuk mengolah data transaksi yang besar menjadi informasi yang lebih terstruktur sehingga dapat dimanfaatkan dalam proses pengambilan keputusan strategis. Dalam industri telekomunikasi, segmentasi pelanggan penting untuk memahami pola pembayaran dan merancang strategi penagihan yang efisien. Penelitian ini menerapkan KDD dengan model RFM dan algoritma *K-Means Clustering* untuk menganalisis pola pembayaran pelanggan *Indibiz* di PT Telkom Indonesia Witel Sumbagsel, dengan tujuan menghasilkan segmentasi yang jelas guna meningkatkan efektivitas pengelolaan dan strategi penagihan pelanggan.

Berbagai implementasi clustering untuk penganalisisan data pernah dikaji oleh S. Sharyanto dan D. Lestari dengan mengintegrasikan algoritma *K-Means* dan model RFM. Dalam penelitian tersebut, validasi kluster dilakukan pada rentang $k=2$ hingga $k=5$ menggunakan *Davies-Bouldin Index* (DBI) [4]. Pendekatan tersebut dapat dijadikan patokan prosedural, namun dengan pengembangan berbeda pada teknik penentuan *cluster* optimal, di mana metode *Elbow* akan diterapkan sebagai pembanding.

Penelitian Carudin menerapkan metode serupa pada data transaksi pelanggan dan membentuk enam kluster, yang optimalisasinya dapat ditingkatkan dengan metode *Elbow* [5]. Selain itu, Zahro dkk. Menganalisis 2.252 transaksi pelanggan ADAPTA.Id, menghasilkan dua kluster utama dengan *Silhouette Score* 0,8511, yang menunjukkan kualitas kluster yang sangat baik dan memberikan *insight* untuk strategi pemasaran [6]. Penelitian Yunita dkk. menggunakan 62.677 transaksi ritel dan memanfaatkan *Elbow Method* serta *Silhouette Analysis* untuk menentukan jumlah kluster optimal, menunjukkan bahwa segmentasi RFM membantu memahami pola pembelian pelanggan dan mendukung strategi *Customer Relationship Management* (CRM) [7].

Untuk menyelesaikan masalah ini, penelitian ini menyarankan untuk mengelompokkan pola perilaku pembayaran pelanggan bisnis dalam kelompok-kelompok yang berbeda menggunakan pendekatan *Knowledge Discovery* yang dikombinasikan dengan pendekatan segmentasi pelanggan berbasis analisis RFM yang didukung oleh algoritma *K-Means*. Pendekatan ini tidak sekadar menghasilkan kluster pelanggan, melainkan juga mengubah *knowledge* yang dapat dipahami dari data historis dan dimanfaatkan oleh unit *Payment Collection* dalam pengambilan keputusan yang lebih efektif. Dengan demikian, diharapkan bahwa penelitian ini tidak hanya memberikan kontribusi teoritis untuk penelitian tentang Manajemen Pengetahuan dan Analisis Data, tetapi juga menawarkan solusi praktis yang dapat diterapkan dalam lingkungan organisasi kontemporer.

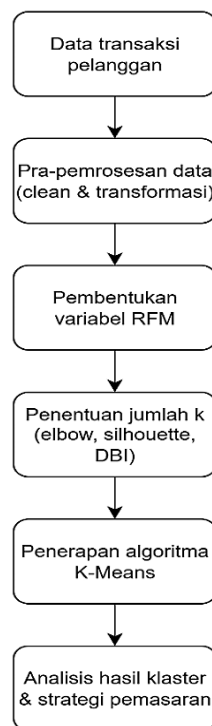
2. METODOLOGI PENELITIAN

Penelitian ini menerapkan metode *Knowledge Discovery in Databases* (KDD) serta algoritma *K-Means Clustering* untuk menyelesaikan permasalahan segmentasi pelanggan dalam industri telekomunikasi. Metode yang diterapkan pada penelitian ini meliputi tahapan pengumpulan data, tahap pra-pengolahan data, penentuan jumlah *cluster* (k) menggunakan metode *elbow* serta *silhouette coefficient*, serta proses menentukan jumlah kluster dengan menerapkan metode *K-Means Clustering*, seperti yang terlihat pada Gambar 1.

2.1 Pengumpulan Data

Proses pengumpulan data adalah tahapan terstruktur yang dilakukan untuk mengumpulkan berbagai informasi yang relevan guna mendukung proses analisis serta pemodelan. Sumber data dalam penelitian ini berupa data sekunder yang diambil dari aplikasi *MyBrains* Telkom. Dataset tersebut berupa informasi pelanggan *Indibiz* pada periode Januari 2025 - Desember 2025. Dataset mencakup atribut *id* pelanggan, jenis produk yang digunakan, kategori pelanggan, waktu riwayat pembayaran dan variabel perilaku

pelanggan lainnya. Data ini berjumlah 46.355, yang mencerminkan perilaku pembayaran pelanggan *Indibiz* berdasarkan indikator variabel di dalam data tersebut. Kuantitas serta variasi data yang diperoleh dinilai telah mencukupi untuk mendukung proses segmentasi pelanggan, khususnya yang berbasis pada pola perilaku serta aspek spasial dan temporal.



Gambar 1. Alur Metode Penelitian

2.2 Pra-pemrosesan Data

Pra-pemrosesan data menjadi tahapan penting dalam menjamin kualitas data agar dapat diolah secara optimal pada tahap analisis sebelum penerapan *clustering* [8]. Tahapan ini mengkaji dan meningkatkan kualitas data guna mengevaluasi konsistensi dan keakuratan hasil analisis data [9]. Proses ini mencakup tiga tahapan, diawali dengan *data cleaning* untuk menangani nilai kosong (*missing values*), menghapus duplikasi transaksi, dan mengeliminasi variabel yang tidak relevan dengan analisis perilaku pembayaran. Langkah berikutnya merupakan transformasi data, yaitu pengolahan lanjutan terhadap data yang selesai melalui tahapan *cleaning*, kemudian dianalisis menggunakan atribut RFM. Pada tahap ini dilakukan penentuan atribut-atribut pada dataset yang dinilai relevan serta memiliki karakteristik yang selaras dengan kebutuhan analisis. Model tersebut mencakup tiga komponen utama, yaitu *Recency* (R), *Frequency* (F), serta *Monetary* (M) [10]. *Recency* menggambarkan jarak waktu sejak transaksi terakhir, *Frequency* menunjukkan jumlah transaksi yang dilakukan pelanggan, sedangkan *Monetary* merepresentasikan total nilai transaksi pelanggan [11].

Selanjutnya atribut RFM akan dinormalisasikan menggunakan metode Indeks Z. Normalisasi Indeks Z (*Zero-Mean Normalization*) digunakan sebagai salah satu teknik dalam proses normalisasi data, di mana nilai setiap atribut disesuaikan pada nilai *mean* dan standar deviasi dari atribut tersebut [12]. Tahapan ini diterapkan pada setiap fitur dalam seluruh sample data [13]. Proses normalisasi dilakukan secara menyeluruh untuk menyeragamkan skala variabel agar tidak ada fitur yang memiliki pengaruh lebih besar dalam analisis akibat perbedaan rentang nilai.

2.3 Elbow Methods

Pada tahap ini setelah dilakukan pra-pemrosesan data, langkah berikutnya adalah menetapkan jumlah *cluster* untuk mendapatkan segmentasi yang efektif dengan menerapkan metode *elbow*. Metode *elbow* berperan dalam mengidentifikasi jumlah kluster optimal dalam proses pengelompokan data dengan mengamati tingkat dari setiap *cluster* yang akan membentuk titik siku pada nilai tertentu. Tujuan dari metode ini untuk menemukan nilai k yang kecil tetapi nilai *insidens* yang rendah [14]. Penentuan jumlah kluster optimal dengan metode *elbow* dilakukan melalui analisis grafik yang membandingkan variasi jumlah kluster (k). Titik terbaik ditunjukkan oleh perubahan yang paling signifikan pada grafik, yang ditandai dengan terbentuknya sudut yang menyerupai siku (*elbow*) pada suatu titik tertentu [15].

2.4 *Silhouette Coefficient*

Penelitian ini menggunakan *Silhouette Coefficient* sebagai metode evaluasi *clustering* untuk menentukan jumlah kluster (K) yang paling cocok. Ini dilakukan dengan membandingkan tingkat kedekatan data terhadap kluster asalnya dan kluster terdekat lainnya [16]. Berikut ini adalah rumus perhitungan *Silhouette Coefficient*.

$$S_i = \frac{b_i - a_i}{\max(a_i - b_i)} \quad (1)$$

dimana a_i merupakan jarak data ke-iii rata-rata dibandingkan dengan seluruh data dalam kluster yang sama. b_i menunjukkan nilai rata-rata jarak terendah data ke-i dalam kluster lain yang paling dekat. Validitas nilai *cluster* berdasarkan Indeks *Silhouette* dengan rentang -1 hingga 1. Nilai yang mendekati 1 merefleksikan pengelompokan dan pemisahan antarkluster yang optimal. Sebaliknya, jika rata-rata nilai *Silhouette Coefficient* mendekati -1, maka hasil pengelompokan tersebut menunjukkan kualitas yang kurang baik [17]. Jumlah kluster terbaik ditentukan berdasarkan nilai K yang menghasilkan skor *Silhouette* paling tinggi.

2.5 *K-Means*

Proses pemodelan kluster dalam penelitian ini dilakukan dengan algoritma *K-Means*, dengan jumlah kluster yang ditetapkan berdasarkan metode *elbow* [18]. Algoritma *K-Means* memiliki beberapa parameter utama yang perlu ditentukan agar proses analisis dapat berjalan secara optimal. Parameter utama yang digunakan dalam teknik ini diantaranya jumlah kluster (K), metode perhitungan jarak dan jumlah iterasi maksimum [19]. Tahapan perhitungan dalam penerapan algoritma *K-Means* dimulai dengan penetapan pusat kluster (y) awal yang dipilih tanpa pola tertentu. Setelah itu, dilakukan pengukuran jarak antara setiap data point (x) terhadap *centroid* menggunakan metode perhitungan *Euclidean Distance* yang bertujuan untuk menghitung tingkat kedekatan antara dua titik data dalam ruang multidimensi [20]. Adapun rumus *Euclidean Distance* ditunjukkan pada Persamaan 2.

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

di mana $d(x, y)$ adalah nilai jarak *Euclidean* yang menggambarkan pemisahan antara dua titik data, x_i dan y_i merupakan komponen ke-i dari dua vektor data dan n menyatakan jumlah dimensi atribut data yang digunakan. Nilai jarak yang diperoleh dari perhitungan ini kemudian digunakan untuk memastikan afiliasi setiap item data dengan *cluster* terdekat dengan *centroid*. Proses ini diulangi dengan menyesuaikan posisi *centroid* hingga kriteria konvergensi terpenuhi, atau terjadi perubahan yang sangat kecil dalam pembentukan kluster. Dengan demikian, algoritma *K-Means* dapat memberikan kluster karakteristik pelanggan yang stabil dan representatif [21].

2.6 *Customer Relationship Management (CRM)*

Sejumlah perusahaan menggunakan *Customer Relationship Management (CRM)* sebagai strategi untuk mengenali pelanggan sebagai upaya meningkatkan kualitas interaksi serta loyalitas pelanggan. CRM merupakan suatu pendekatan strategi bisnis yang berorientasi pada pelanggan, yang diterapkan perusahaan untuk meningkatkan kesetiaan pelanggan serta menjalin hubungan berkelanjutan yang memberikan nilai tambah [22]. Implementasi CRM mendukung organisasi bisnis dalam mengoptimalkan nilai tambah pelanggan sekaligus mendorong pencapaian keuntungan yang lebih tinggi [23]. Segmentasi pelanggan merupakan salah satu tahapan fundamental dalam penerapan CRM, yang dilakukan dengan mengelompokkan konsumen ke dalam beberapa kategori berdasarkan kesamaan karakteristik tertentu [24].

2.7 *Model Recency, Frequency, Monetary (RFM)*

Pendekatan analisis RFM awal mula dikemukakan oleh Hughes pada tahun 1994 sebagai suatu pendekatan untuk menilai nilai pelanggan yang kemudian digunakan sebagai dasar dalam melakukan segmentasi pelanggan [25]. Model RFM dimanfaatkan untuk mengidentifikasi pelanggan bernilai tinggi dari data transaksi dalam jumlah besar, sehingga membantu dalam strategi pemasaran dan penagihan. Teknik RFM sangat mengandalkan data transaksi yang dihasilkan oleh aktivitas konsumen [17]. Variabel RFM digunakan untuk mengidentifikasi nilai pelanggan melalui perilaku dan riwayat pembayaran pelanggan di tiga dimensi utama. *Recency*, *Frequency*, *Monetary* dimanfaatkan untuk mengelompokkan pelanggan berdasarkan waktu terakhir terjadinya transaksi, tingkat intensitas pembelian, serta kontribusi nilai finansial yang dihasilkan dari aktivitas tersebut. Dalam Konteks penelitian ini, pendekatan RFM digunakan untuk mengidentifikasi karakteristik perilaku pelanggan *Indibiz*. Melalui pendekatan ini, perusahaan dapat mengidentifikasi kelompok pelanggan berdasarkan tingkat aktivitas pembayaran, sehingga dapat mendukung penyusunan strategi penagihan yang lebih tepat sasaran dan efisien [26].

3. HASIL DAN PEMBAHASAN

3.1. Data Selection

Data diperoleh melalui proses penarikan data dari sistem informasi *Mybrains* PT Telkom Witel Sumbagsel. Data pelanggan yang dimanfaatkan ini adalah dataset riwayat pembayaran konsumen mulai dari periode bulan Januari 2025 - Desember 2025 dengan total 46.355 transaksi. Data mengandung beberapa variabel, tetapi tidak semua dari mereka akan digunakan saat melakukan segmentasi pelanggan di masa mendatang. Dataset penelitian ini disajikan dalam Tabel 1.

Tabel 1. *Data Processing*

SND	Produk	Desc Newbill	Saldo	Tanggal Call	Status
*****81216	Wifi Manage Service	PRANPC	244200	2/5/2025	TRACKING FULLFILMENT
*****54319	Wifi Manage Service	Eksisting	266400	2/5/2025	NOMOR TIDAK ADA
*****33930	Internet	Eksisting	28860	2/5/2025	LUNAS
*****10138	Internet	Eksisting	576090	2/5/2025	NOMOR TIDAK ADA
*****61421	Internet	Eksisting	429570	2/5/2025	LUNAS

Berdasarkan Tabel 1, data yang dimanfaatkan dalam penelitian ini meliputi beberapa variabel yang merepresentasikan karakteristik pelanggan dan riwayat transaksinya. Variabel SND berfungsi sebagai identitas unik pelanggan, Produk menunjukkan jenis layanan yang digunakan, dan Desc Newbill menggambarkan kategori pelanggan. Variabel Saldo merepresentasikan nilai tagihan, Tanggal Call menunjukkan waktu pencatatan interaksi, sedangkan Status menggambarkan hasil komunikasi atau kondisi pembayaran pelanggan.

3.2. Data Preprocessing

Selanjutnya, tahap preprocessing data terdiri dari data cleaning, di mana setiap variabel yang dianggap tidak relevan akan dihilangkan. Pada penelitian ini hanya variabel yang dinilai penting yang dipertahankan, sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. *Data Preprocessing*

Tanggal Call	SND	Saldo
12/16/2025	*****09279	151331
12/16/2025	*****09309	140600
12/16/2025	*****09314	199800
12/16/2025	*****08306	429570
12/16/2025	*****16538	508103
12/16/2025	*****08657	411626
12/16/2025	*****16617	628076
12/16/2025	*****11676	396492
12/16/2025	*****08658	314896

Berdasarkan Tabel 2, tahap *data preprocessing* dilakukan dengan menghilangkan variabel yang tidak relevan dan hanya mempertahankan atribut yang dimanfaatkan dalam analisis, yaitu Tanggal Call, SND, dan Saldo.

3.3. Data Transformation

Pada tahap ini, variabel yang telah melalui proses *preprocessing* selanjutnya diolah ke dalam model RFM (*Recency, Frequency, Monetary*). Nilai *recency* berdasarkan selisih periode waktu sejak transaksi terakhir, *Frequency* berdasarkan banyaknya transaksi yang dilakukan pelanggan, dan *monetary* berdasarkan jumlah transaksi yang dilakukan konsumen. Proses transformasi ini menghasilkan nilai atribut *Recency, Frequency, dan Monetary* yang ditunjukkan dalam Tabel 3.

Tabel 3. Hasil RFM

SND	Recency	Frequency	Monetary
*****01013	153	1	260850
*****01060	156	1	63936
*****01062	156	1	72927
*****01065	156	1	63936

SND	Recency	Frequency	Monetary
*****01067	153	1	2640033

Berdasarkan Tabel 3, setiap pelanggan memiliki nilai Recency, Frequency dan Monetary yang berbeda-beda, yang mencerminkan pola aktivitas dan kontribusi transaksi masing-masing pelanggan.

Setelah data ditransformasikan ke dalam model RFM, data transaksi pelanggan kemudian diproses menjadi skor RFM dengan cara menghitung serta memberikan bobot pada setiap variabel. Hasil perhitungan skor RFM ditunjukkan pada Tabel 4.

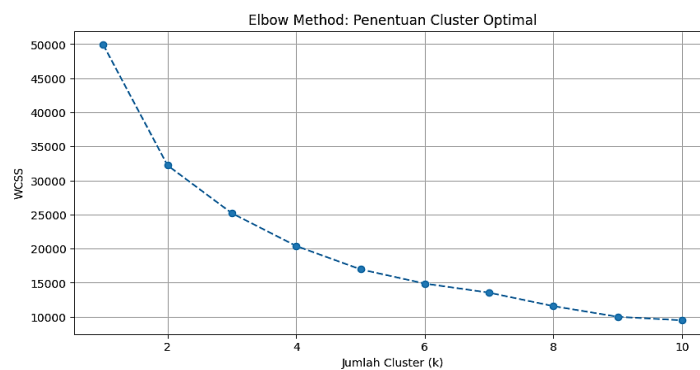
Tabel 4. Skor RFM

SND	Recency	Frequency	Monetary	RFM Score
*****01013	153	1	260850	6
*****01060	156	1	63936	5
*****01062	156	1	72927	5
*****01065	156	1	63936	5
*****01067	153	1	2640033	6

Berdasarkan Tabel 4, setiap pelanggan memiliki nilai RFM score yang berbeda-beda sesuai dengan pola kombinasi nilai *Recency*, *Frequency*, dan *Monetary* yang dimilikinya. Skor yang lebih tinggi menunjukkan pelanggan dengan tingkat aktivitas dan kontribusi yang relatif baik.

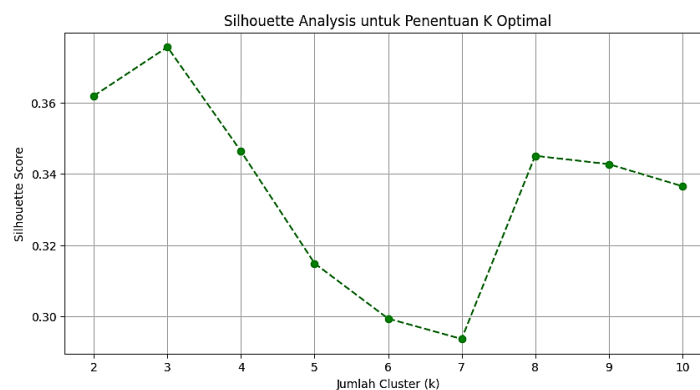
3.4. Data mining

Selanjutnya, metode *elbow* diterapkan untuk menetapkan jumlah kluster (k) yang ideal.



Gambar 2. Grafik Elbow

Dapat dilihat pada Gambar 2, grafik metode *elbow* yang dihasilkan, penentuan jumlah cluster (k) terbaik untuk penelitian di unit *Payment Collection* PT Telkom Witel Sumbagsel adalah $k=4$. Setelah itu dilakukan evaluasi *Silhouette Coefficient* untuk mengukur skor optimal dari nilai K yang ditentukan.



Gambar 3. Grafik Silhouette Analysis

Berdasarkan Gambar 3, hasil pengujian menggunakan *Silhouette Coefficient*, model *clustering* dengan $k=4$ menghasilkan skor sebesar 0,3463. Hasil ini menunjukkan bahwa kualitas pengelompokan lebih baik jika dibandingkan dengan penelitian terdahulu yang hanya mencapai skor 0,2850. Hal ini memperlihatkan

bahwa algoritma *K-Means* dapat dengan mudah menemukan segmen pelanggan *Indibiz* berdasarkan pola perilaku pembayaran yang berguna dan valid. Tabel 5 di bawah ini menunjukkan hasil clustering.

Tabel 5. Hasil *Clustering*

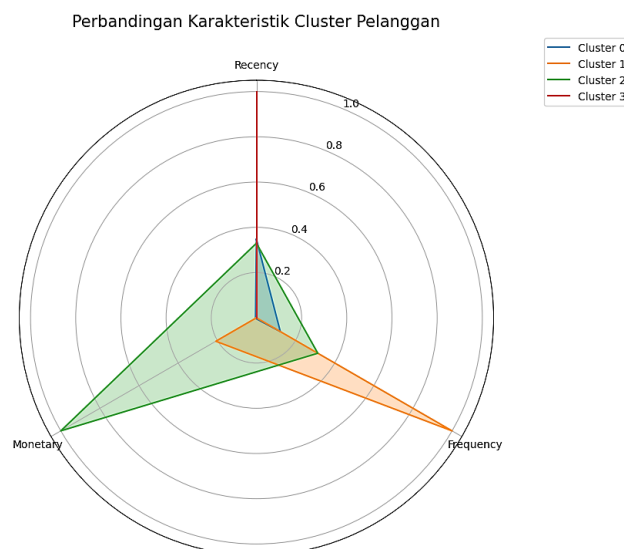
<i>Cluster</i>	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>	Jumlah Pelanggan
0	146.675524	1.936979	261239	9711
1	73.722265	5.441483	432891	3939
2	143.112597	2.701619	1104244	1421
3	284.297315	1.449488	257040	1564

Tabel 5 menunjukkan empat *cluster* yang terbentuk beserta nilai *Recency*, *Frequency*, *Monetary* dan jumlah pelanggan pada masing-masing *cluster*.

3.5. Evaluasi dan Interpretasi

Silhouette Coefficient digunakan untuk mengevaluasi guna memastikan validitas pengelompokan. Model $k=4$ mendapatkan skor sebesar 0.3463. Nilai positif ini mengindikasikan bahwa data telah terkelompok dengan struktur yang cukup jelas, dimana jarak keterkaitan data dalam satu *cluster* lebih dekat daripada dengan jarak ke *cluster* lain. Hal ini mengonfirmasi bahwa segmentasi yang dilakukan mampu membedakan perilaku pembayaran pelanggan *Indibiz* secara efektif.

Setelah menentukan jumlah *cluster* optimal sebanyak $k=4$, langkah selanjutnya adalah menganalisis karakteristik perilaku pembayaran dari masing-masing segmen pelanggan. Untuk memberikan gambaran komprehensif mengenai perbedaan pola tersebut, dilakukan visualisasi perbandingan antarklaster menggunakan *radar chart*. Grafik ini menggambarkan *mean* nilai dari atribut *Recency*, *Frequency*, dan *Monetary* yang telah dinormalisasi, sehingga profil unik dari setiap kelompok dapat diidentifikasi secara objektif sebagaimana ditunjukkan pada Gambar 5.



Gambar 5. Perbandingan Karakteristik RFM Tiap Klaster Pelanggan

Grafik di atas mengilustrasikan perbedaan karakteristik pelanggan di antara tiga kategori dari atribut *Recency*, *Frequency*, dan *Monetary* yang telah dinormalisasi. Visualisasi ini menunjukkan bahwa setiap kelas memiliki profil Perilaku yang unik. Frekuensi yang lebih tinggi menunjukkan tingkat kepuasan pelanggan yang lebih tinggi dalam interaksi atau transaksi, sementara frekuensi moneter yang lebih besar memperlihatkan kontribusi finansial yang lebih tinggi bagi perusahaan. Analisis segmentasi pelanggan ini konsisten dengan literatur terkini tentang RFM dan pengelompokan *K-Means* sebagai sarana untuk memahami perilaku konsumen dan meningkatkan efektivitas strategi pemasaran [7].

Dari grafik terlihat bahwa Cluster 3 memiliki nilai tertinggi pada dimensi *Frequency*, yang menunjukkan kelompok pelanggan dengan tingkat interaksi paling intens terhadap layanan. Sementara itu, Cluster 2 menunjukkan nilai paling dominan pada aspek *Monetary*, menandakan bahwa segmen ini memberikan kontribusi finansial lebih besar dibanding klaster lainnya. Adapun Cluster 0 dan Cluster 1 memperlihatkan nilai yang cenderung moderat pada ketiga aspek RFM, sehingga mencerminkan pola perilaku yang lebih stabil atau berada pada tahap transisi. Pola ini menggambarkan bahwa setiap klaster dapat

diarahkan pada strategi bisnis yang berbeda guna meningkatkan retensi serta nilai pelanggan secara keseluruhan. Berikut adalah interpretasi lebih lanjut dari masing-masing klaster:

1. Klaster 0: *Potential Risk* (*Recency* sedang, *Frequency* sedang, *Monetary* sedang-rendah)
 - a. Interpretasi: Merupakan kelompok pelanggan mayoritas dengan pola perilaku pembayaran yang berada pada tingkat rata-rata di seluruh dimensi RFM.
 - b. Strategi: Optimalisasi pengingat otomatis (*automated reminders*) berbasis WhatsApp/SMS dan pemberian program poin insentif bagi pelanggan yang membayar tepat waktu.
2. Klaster 1: *Chronic Delinquents* (*Recency* berada pada tingkat rendah, *Frequency* pada tingkat paling tinggi, dan *Monetary* berada pada kategori sedang)
 - a. Interpretasi: Pelanggan yang paling rutin masuk dalam daftar penagihan dengan intensitas interaksi yang sangat baru.
 - b. Strategi: Edukasi penggunaan sistem pembayaran otomatis atau *auto-debet* serta penerapan sanksi isolir bertahap untuk mendorong kedisiplinan pembayaran.
3. Klaster 2: *High-Value At-Risk* (*Recency* sedang, *Frequency* sedang, *Monetary* tertinggi)
 - a. Interpretasi: Kelompok pelanggan dengan nilai tunggakan saldo terbesar (rata-rata > Rp1,1 Juta) yang berdampak signifikan pada pendapatan perusahaan.
 - b. Strategi: Pendekatan personal melalui kunjungan langsung (*door-to-door*) atau penawaran negosiasi termin pembayaran khusus guna mengamankan piutang bernilai besar.
4. Klaster 3: *Low Risk/Dormant* (*Recency* tertinggi, *Frequency* terendah, *Monetary* terendah)
 - a. Interpretasi: Pelanggan yang sangat jarang menunggak dan secara historis sudah sangat lama tidak masuk dalam sistem penagihan.
 - b. Strategi: Pengawasan minimal untuk efisiensi biaya operasional unit serta pemberian pesan apresiasi loyalitas guna menjaga kepuasan pelanggan jangka panjang.

4. KESIMPULAN

Berdasarkan evaluasi menggunakan *Silhouette Coefficient* serta hasil *profiling cluster*, penggunaan model RFM dan algoritma *K-Means* terbukti mampu bekerja secara efektif. dalam mentransformasi data transaksi pelanggan menjadi informasi strategis pada unit *Payment Collection* PT Telkom Indonesia Witel Sumbagsel. Analisis terhadap 46.355 data transaksi menghasilkan jumlah *cluster* optimal sebesar $k = 4$ dengan nilai *Silhouette Coefficient* 0,3463, yang mengindikasikan struktur segmentasi yang valid dan mampu merepresentasikan perbedaan perilaku pembayaran pelanggan secara jelas. Segmentasi yang dihasilkan mengidentifikasi empat kelompok pelanggan dengan karakteristik pembayaran yang *distinct*, sehingga memungkinkan penerapan strategi penagihan yang lebih terarah, adaptif, dan berbasis data. Temuan ini menegaskan bahwa integrasi RFM dan *K-Means* memberikan landasan analitis yang kuat dalam mendukung pengambilan keputusan serta peningkatan efisiensi pengelolaan pelanggan pada industri telekomunikasi. Secara praktis, studi ini memberikan manfaat langsung bagi optimalisasi *Customer Relationship Management* (CRM) di PT Telkom Witel Sumbagsel. Dengan pelanggan yang terbagi ke dalam 4 empat klaster profil pelanggan, unit *Payment Collection* dapat mengalokasikan sumber daya penagihan secara lebih efisien. Untuk penelitian selanjutnya, disarankan membandingkan *K-Means* dengan algoritma alternatif seperti *K-Medoids* atau *DBSCAN* untuk mencapai pengelolaan data outlier yang lebih efisien dalam kumpulan data transaksi yang luas.

REFERENSI

- [1] R. D. Nugraha, D. D. Adelia, dan D. Rivaldi, "Segmentasi Pelanggan Retail Berbasis Perilaku menggunakan Algoritma *K-Means Clustering*," *Digit. Transform. Technol.*, vol. 5, no. 2, hlm. 141–148, Okt 2025, doi: 10.47709/digitech.v5i2.6340.
- [2] N. H. Harani, C. Prianto, dan F. A. Nugraha, "Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma *K-Means* Berbasis Python," *J. Manaj. Inform. JAMIKA*, vol. 10, no. 2, hlm. 133–146, Okt 2020, doi: 10.34010/jamika.v10i2.2683.
- [3] F. Arifianto, J. Hasudungan, A. Muzaky, dan H. T. Y. Achsan, "Segmentasi Pelanggan Berdasarkan *Recency*, *Frequency*, dan *Monetary* dengan *K-Means Clustering*: Studi Kasus Toko Pakaian Almost Famous," *J. Teknol. Inform. Dan Komput.*, vol. 10, no. 1, hlm. 122–135, Mar 2024, doi: 10.37012/jtik.v10i1.2096.
- [4] S. Sharyanto dan D. Lestari, "Penerapan *Data mining* Untuk Menentukan Segmentasi Pelanggan Dengan Menggunakan Algoritma *K-Means* dan Model RFM Pada E-Commerce," *JURIKOM J. Ris. Komput.*, vol. 9, no. 4, hlm. 866, Agu 2022, doi: 10.30865/jurikom.v9i4.4525.

- [5] Carudin, "Pemanfaatan Data Transaksi untuk Dasar membangun Strategi berdasarkan Karakteristik Pelanggan dengan Algoritma *K-Means Clustering* dan Model RFM," *J. Teknol. Terpadu*, vol. 7, no. 1, hlm. 7–14, Jul 2021, doi: 10.54914/jtt.v7i1.318.
- [6] N. Zahro, N. A. Maori, dan G. W. Nyipto Wibowo, "Integration of RFM Method and *K-Means Clustering* for Customer Segmentation Effectiveness," *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 5, no. 1, hlm. 12–21, Feb 2025, doi: 10.20895/dinda.v5i1.1649.
- [7] I. Yunita, P. R. Ali, M. A. Kartawidjaja, dan R. Sukwadi, "Segmentasi Pelanggan Menggunakan *K-Means Clustering*: Menganalisis Metrik RFM untuk Strategi Pemasaran," *J. Media Tek. Dan Sist. Ind.*, vol. 9, no. 1, hlm. 58, Mar 2025, doi: 10.35194/jmtsi.v9i1.4452.
- [8] F. Putra, H. F. Tahiyat, R. M. Ihsan, R. Rahmadden, dan L. Efrizoni, "Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia: Application of K-Nearest Neighbor Algorithm Using Wrapper as Preprocessing for Determination of Human Weight Information," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, hlm. 273–281, Jan 2024, doi: 10.57152/malcom.v4i1.1085.
- [9] C. Fan, M. Chen, X. Wang, J. Wang, dan B. Huang, "A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data," *Front. Energy Res.*, vol. 9, hlm. 652801, Mar 2021, doi: 10.3389/fenrg.2021.652801.
- [10] V. Hermawati dan R. Sulaiman, "Penentuan Segmentasi Pelanggan E-Commerce Menggunakan Fuzzy C-Means Dan Model Fuzzy Rfm," *MATHunesa J. Ilm. Mat.*, vol. 9, no. 1, hlm. 76–88, Jan 2021, doi: 10.26740/mathunesa.v9n1.p76-88.
- [11] K. Chouaten, C. Rodriguez Rivero, F. Nack, dan M. Reckers, "Unlocking high-value football fans: unsupervised machine learning for customer segmentation and lifetime value," *Front. Sports Act. Living*, vol. 6, hlm. 1362489, Agu 2024, doi: 10.3389/fspor.2024.1362489.
- [12] "Identifikasi Predikat Hasil Pengelompokan Data Kualitas Udara dengan Menggunakan Affinity Propagation dan Silhouette Coefficient," *InfoTekJar J. Nas. Inform. Dan Teknol. Jar.*.
- [13] A. Dursun dan M. Caber, "Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis," *Tour. Manag. Perspect.*, vol. 18, hlm. 153–160, Apr 2016, doi: 10.1016/j.tmp.2016.03.001.
- [14] I. T. Utami, F. Suryaningrum, dan D. Ispriyanti, "K-Means Cluster Count Optimization With Silhouette Index Validation And Davies Bouldin Index (Case Study: Coverage Of Pregnant Women, Childbirth, And Postpartum Health Services In Indonesia In 2020)," *BAREKENG J. Ilmu Mat. Dan Terap.*, vol. 17, no. 2, hlm. 0707–0716, Jun 2023, doi: 10.30598/barekengvol17iss2pp0707-0716.
- [15] D. Tivano, I. Ruslianto, dan D. Prawira, "Sistem Segmentasi Pelanggan Potensial Menggunakan Metode K-Means Dan Analisis RFM," *Coding J. Komput. Dan Apl.*, vol. 11, no. 3, hlm. 325, Des 2023, doi: 10.26418/coding.v11i03.53238.
- [16] Siti Nurhaliza dan Tessy Octavia Mukhti, "Clustering Regions in West Sumatera Based on the Special Protection Index for Children Using *K-Means Clustering* with Silhouette Coefficient," *UNP J. Stat. Data Sci.*, vol. 3, no. 1, hlm. 123–129, Feb 2025, doi: 10.24036/ujsds/vol3-iss1/356.
- [17] R. A. Farissa, R. Mayasari, dan Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsambung," *J. Appl. Inform. Comput.*, vol. 5, no. 2, hlm. 109–116, Okt 2021, doi: 10.30871/jaic.v5i1.3237.
- [18] N. Ahsina, F. Fatimah, dan F. Rachmawati, "Analisis Segmentasi Pelanggan Bank Berdasarkan Pengambilan Kredit Dengan Menggunakan Metode *K-Means Clustering*," *J. Ilm. Teknol. Infomasi Terap.*, vol. 8, no. 3, Agu 2022, doi: 10.33197/jitter.vol8.iss3.2022.883.
- [19] M. Ramadhan Putra Pratama dkk., "Perbandingan Algoritma K-Means, K-Medoid, Dan DbSCAN Untuk Clustering Kualitas Hidup Indonesia Dalam Perspektif Knowledge Management Dan Data Discovery," *JATI J. Mhs. Tek. Inform.*, vol. 9, no. 4, hlm. 5903–5910, Mei 2025, doi: 10.36040/jati.v9i4.13916.
- [20] N. Puspitasari, H. Havaluddin, dan F. U. J. Helmi Puadi, "Klasterisasi Wilayah Penghasil Tanaman Lada Menggunakan Algoritma K-Means," *Indones. J. Comput. Sci.*, vol. 11, no. 3, Des 2022, doi: 10.33022/ijcs.v11i3.3104.
- [21] A. Tasya, G. Kholijah, dan S. Sarmada, "The Use of Elbow K-Means And K-Medoids in the Grouping of Provinces in Indonesia Based on the Indicators of the Effectiveness of the Authentication Taspen Application With DBI and Silhouette Coefficients," *ALACRITY J. Educ.*, hlm. 1129–1143, Okt 2025, doi: 10.52121/alacrity.v5i3.891.
- [22] S.-I. Wu dan C.-L. Lu, "The relationship between CRM, RM, and business performance: A study of the hotel industry in Taiwan," *Int. J. Hosp. Manag.*, vol. 31, no. 1, hlm. 276–285, Mar 2012, doi: 10.1016/j.ijhm.2011.06.012.
- [23] R. Thakur dan L. Workman, "Customer portfolio management (CPM) for improved customer relationship management (CRM): Are your customers platinum, gold, silver, or bronze?," *J. Bus. Res.*, vol. 69, no. 10, hlm. 4095–4102, Okt 2016, doi: 10.1016/j.jbusres.2016.03.042.

- [24] Y.-S. Chen, C.-H. Cheng, C.-J. Lai, C.-Y. Hsu, dan H.-J. Syu, "Identifying patients in target customer segments using a two-stage clustering-classification approach: A hospital-based assessment," *Comput. Biol. Med.*, vol. 42, no. 2, hlm. 213–221, Feb 2012, doi: 10.1016/j.compbiomed.2011.11.010.
- [25] A. M. Hughes, *Strategic database marketing: masterplan for starting and managing a profitable, customer-based marketing program*. Chicago, Ill, Cambridge: Probus Publishing, 1994.
- [26] G. W. A. Dwi dan A. D. Wowor, "Segmentasi pelanggan pada penjualan perusahaan air minum XYZ dengan analisis RFM dan *K-Means Clustering*," *IT-Explore J. Penerapan Teknol. Inf. Dan Komun.*, vol. 5, no. 1, hlm. 75–85, Feb 2026, doi: 10.24246/itexplore.v5i1.2026.pp75-85.