

ANALISIS SEGMENTASI PELANGGAN E-COMMERCE INDIA BERBASIS RFM DAN PROFITABILITAS MENGGUNAKAN K-MEANS

Deny Saputra, Untung Subagyo

Teknik Informatika, Sekolah Tinggi Manajemen dan Ilmu Komputer El Rahma Yogyakarta
Jalan Sisimangaraja No. 76, Yogyakarta, DI Yogyakarta
Dhenysap77@gmail.com

ABSTRAK

Perkembangan industri e-commerce mendorong perusahaan untuk memahami perilaku pelanggan guna meningkatkan efektivitas strategi pemasaran. Namun, keberagaman karakteristik pelanggan menyebabkan perusahaan kesulitan mengidentifikasi pelanggan yang berkontribusi terhadap profitabilitas. Penelitian ini bertujuan melakukan segmentasi pelanggan menggunakan metode Recency, Frequency, Monetary (RFM) dan analisis profitabilitas dengan algoritma K-Means. Metode penelitian menggunakan pendekatan Knowledge Discovery in Databases (KDD) yang meliputi data selection, preprocessing, transformation, clustering, dan evaluation. Dataset yang digunakan merupakan data transaksi Indian e-commerce periode 2023–2025 sebanyak 30.600 data. Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow dan evaluasi menggunakan Silhouette Score. Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah tiga cluster dengan nilai Silhouette Score sebesar 0,4747. Cluster yang dihasilkan terdiri dari Low Loyalty sebanyak 13.707 pelanggan (44,8%), Medium Loyalty sebanyak 13.634 pelanggan (44,5%), dan High Loyalty sebanyak 3.259 pelanggan (10,7%). Analisis profitabilitas menunjukkan kontribusi revenue masing-masing cluster sebesar 33%, 36%, dan 31%. Hasil segmentasi dapat digunakan untuk menentukan strategi pemasaran, retensi pelanggan, dan peningkatan profitabilitas perusahaan.

Kata kunci: *E-commerce, Segmentasi Pelanggan, RFM, Profitabilitas, K-Means Clustering, KDD.*

1. PENDAHULUAN

Perkembangan teknologi digital mendorong pertumbuhan industri e-commerce yang menghasilkan data transaksi dalam jumlah besar dan kompleks. Data tersebut mengandung informasi penting yang dapat dimanfaatkan perusahaan untuk memahami perilaku pelanggan dan mendukung pengambilan keputusan strategis [1]. Namun, tingginya heterogenitas karakteristik pelanggan menyebabkan perusahaan mengalami kesulitan dalam mengidentifikasi pelanggan yang memberikan kontribusi terbesar terhadap pendapatan dan keuntungan perusahaan [2].

Dalam mengatasi permasalahan tersebut, data mining menjadi salah satu pendekatan yang efektif untuk mengekstraksi pola dan pengetahuan dari data dalam jumlah besar. Salah satu teknik yang banyak digunakan adalah algoritma K-Means clustering karena mampu mengelompokkan pelanggan berdasarkan kemiripan karakteristik tertentu [2], [3]. Selain itu, model Recency, Frequency, dan Monetary (RFM) juga banyak diterapkan untuk menganalisis perilaku pelanggan berdasarkan waktu transaksi terakhir, frekuensi pembelian, dan total nilai transaksi pelanggan [4], [5].

Meskipun sebagian besar penelitian sebelumnya berfokus pada penggunaan metode RFM tanpa mempertimbangkan aspek profitabilitas secara langsung, pendekatan tersebut masih memiliki keterbatasan [11]. Padahal pelanggan dengan frekuensi transaksi tinggi belum tentu memberikan kontribusi keuntungan yang optimal untuk perusahaan. Penelitian sebelumnya menunjukkan bahwa segmentasi berbasis RFM efektif dalam pengelompokan pelanggan

berdasarkan tingkat Loyalty, namun belum mengintegrasikan analisis profitabilitas, sehingga belum mampu merepresentasikan kontribusi nilai ekonomi pelanggan secara komprehensif [4], [6]. Hal ini menunjukkan research gap dalam pengembangan segmentasi pelanggan yang lebih komprehensif.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan model segmentasi pelanggan dengan mengintegrasikan metode RFM dan variabel profitabilitas menggunakan algoritma K-Means clustering pada dataset Indian e-commerce periode 2023–2025. Proses analisis dilakukan menggunakan pendekatan Knowledge Discovery in Databases (KDD), dengan penentuan jumlah cluster menggunakan metode Elbow serta evaluasi kualitas cluster menggunakan Silhouette Score. Integrasi variabel profitabilitas diharapkan mampu menghasilkan segmentasi pelanggan yang lebih akurat dalam menggambarkan kontribusi pelanggan terhadap pertumbuhan perusahaan.

Kontribusi utama adalah menghasilkan model segmentasi pelanggan berbasis RFM dan profitabilitas yang mampu meningkatkan kualitas serta memberikan dasar yang lebih kuat dalam penyusunan strategi pemasaran, khususnya meningkatkan retensi pelanggan, efektivitas penargetan, dan efisiensi promosi pada industri e-commerce [3].

2. TINJAUAN PUSTAKA

2.1. Penelitian terdahulu

Penelitian mengenai segmentasi pelanggan menggunakan metode *Recency, Frequency, Monetary* (RFM) dan algoritma *K-Means Clustering* telah

banyak diterapkan dalam berbagai studi, khususnya pada bidang e-commerce dan retail.

Arifianto dkk. [1] melakukan segmentasi pelanggan menggunakan pendekatan RFM dan K-Means pada studi kasus toko pakaian. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengelompokkan pelanggan berdasarkan tingkat loyalitas sehingga dapat digunakan untuk mendukung strategi pemasaran. Namun, penelitian tersebut hanya berfokus pada perilaku transaksi tanpa mempertimbangkan aspek profitabilitas pelanggan.

Imanuel dan Alfian [3] menerapkan visualisasi segmentasi pelanggan berbasis RFM dengan K-Means pada retail online. Penelitian tersebut mampu memberikan gambaran karakteristik setiap cluster pelanggan melalui visualisasi data. Akan tetapi, penelitian ini belum memasukkan variabel kontribusi keuntungan pelanggan dalam proses clustering.

Febriana dan Rosadi [4] menggunakan K-Means Clustering untuk segmentasi pelanggan e-commerce berdasarkan model RFM. Hasil penelitian menunjukkan bahwa metode tersebut efektif dalam mengelompokkan pelanggan berdasarkan tingkat loyalitas. Namun, penelitian belum mempertimbangkan nilai profit pelanggan sebagai variabel analisis.

Nugraha dkk. [6] melakukan segmentasi pelanggan berbasis RFM menggunakan K-Means Clustering. Hasil penelitian menunjukkan bahwa metode RFM efektif dalam mengidentifikasi pola pelanggan berdasarkan aktivitas transaksi. Namun, pendekatan tersebut masih terbatas karena belum mengintegrasikan aspek profitabilitas pelanggan.

Berdasarkan beberapa penelitian tersebut, dapat disimpulkan bahwa metode RFM dan K-Means Clustering efektif dalam melakukan segmentasi pelanggan berdasarkan perilaku transaksi dan loyalitas. Namun, sebagian besar penelitian masih memiliki keterbatasan karena belum mengintegrasikan variabel profitabilitas pelanggan, sehingga belum mampu menggambarkan kontribusi ekonomi pelanggan secara komprehensif.

Oleh karena itu, penelitian ini mengusulkan model segmentasi pelanggan berbasis integrasi RFM dan profitabilitas menggunakan algoritma K-Means Clustering. Pendekatan ini diharapkan mampu menghasilkan segmentasi pelanggan yang lebih akurat dengan mempertimbangkan tidak hanya perilaku transaksi, tetapi juga kontribusi keuntungan pelanggan terhadap perusahaan.

2.2. Data Mining

Data mining merupakan proses analisis yang dilakukan secara interaktif pada basis data berukuran besar. Data mining juga bisa disebut sebagai serangkaian proses mengekstraksi atau menambang pengetahuan melalui data warehouse atau mengeksplor dan mencari nilai yang terdapat pada data tersebut.

Istilah data mining disebut juga sebagai Knowledge discovery in Databases (KDD) sering digunakan untuk mengklasifikasi data pelanggan dalam menentukan strategi pemasaran [2].

2.3. Algoritma K-Means

Algoritma K-Means dikenal sebagai metode pengelompokan yang sederhana, efisien, dan tangguh. Keunggulan tersebut menjadikannya pilihan utama dalam berbagai konteks, seperti segmentasi pelanggan, segmentasi citra, dan pengelompokan [7].

K-Means merupakan salah satu algoritma clustering yang digunakan untuk pengelompokan data ke dalam beberapa cluster berdasarkan tingkat kemiripan karakteristik [6]. Algoritma ini bekerja dengan cara meminimalkan jarak antara data dengan pusat (centroid), sehingga data dalam satu cluster memiliki tingkat kemiripan yang tinggi [8].

Metode ini banyak digunakan dalam berbagai bidang, termasuk segmentasi pelanggan, karena memiliki keunggulan dalam hal kesederhanaan, efisiensi, dan kemampuan dalam mengolah data berukuran besar [9].

Proses kerja algoritma K-Means meliputi beberapa langkah yaitu

- Menentukan jumlah cluster (k).
- Menginisialisasi centroid secara acak.
- Menghitung jarak setiap data ke centroid terdekat. Berikut persamaan Euclidean Distance yang ditunjukkan pada persamaan:

$$d(x_i, u_j) = \sqrt{\sum (x_i - u_j)^2} \quad (1) \quad [10]$$

x_i : Data Kriteria
 u_j : Centroid Pada Cluster Le-J

$d(x_i, u_j)$: Klasifikasi Kedekatan Centroid

- Mengelompokkan data berdasarkan centroid
- Memperbarui posisi centroid yang dihubungkan dengan rumus[3] :

$$u_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in s_j} x_j \quad (2) \quad [10]$$

$u_j(t+1)$: Centroid Baru
 $ke(t+1)N_{sj}$: Banyak data Cluster s_j
 x : Data dalam Cluster

- Mengulangi langkah hingga centroid stabil

K-Means banyak digunakan dalam berbagai bidang seperti segmentasi pelanggan, dan analisis pola data. Membantu perusahaan mendapatkan strategi pemasaran yang lebih tepat.[1]

2.4. Model RFM

Recency, Frequency, dan Monetary adalah teknik yang sering digunakan dalam mengevaluasi pelanggan berdasarkan karakteristik pembeli. RFM digunakan untuk mengelompokkan pelanggan yang melibatkan sejumlah pelanggan pembelian ritel [10]. Metode ini mengelompokkan pelanggan berdasarkan tiga indikator sebagai berikut:

a. Recency

Menunjukkan kebaruan aktivitas atau pembelian pelanggan. Dalam konteks data ini, Recency dihitung berdasarkan tanggal terakhir kunjungan pelanggan dan tanggal terbaru yang terdapat dalam dataset sebagai titik acuan untuk mengukur tingkat kebaruan [1].

b. Frequency

Frequency untuk mengukur nilai pelanggan berdasarkan seberapa sering pelanggan melakukan transaksi dalam satu periode,

c. Monetary

Mengukur total pengeluaran pelanggan dalam periode tertentu. Data yang dihitung mencakup jumlah nilai yang dihasilkan oleh pelanggan selama berbelanja pada tempat tersebut.

RFM banyak digunakan dalam segmentasi pelanggan karena mampu memberikan gambaran tingkat loyalitas dan nilai pelanggan [3].

2.5. Profitabilitas Pelanggan

Profitabilitas pelanggan merupakan ukuran yang digunakan untuk mengetahui kontribusi keuntungan yang dihasilkan oleh pelanggan dalam suatu periode tertentu. Variabel ini penting karena pelanggan dengan aktivitas transaksi tinggi belum tentu memberikan keuntungan yang besar bagi perusahaan.

Dengan mengintegrasikan profitabilitas ke dalam model RFM, segmentasi pelanggan menjadi lebih komprehensif karena tidak hanya mempertimbangkan perilaku transaksi, tetapi kontribusi nilai ekonomi pelanggan juga diperlukan [2].

2.6. Keterkaitan RFM, Profitabilitas, dan K-Means

Dalam penelitian ini, model RFM dan variabel profitabilitas digunakan sebagai atribut dalam proses clustering menggunakan algoritma K-Means. Pendekatan ini memungkinkan pengelompokan pelanggan berdasarkan kemiripan perilaku transaksi dan kontribusi keuntungan secara simultan [11].

Dengan demikian, hasil clustering yang diperoleh tidak hanya menggambarkan tingkat loyalitas pelanggan, tetapi juga memberikan informasi yang lebih akurat dalam mendukung pengambilan keputusan strategi pemasaran, khususnya dalam penentuan strategi dan peningkatan retensi pelanggan.

3. METODE PENELITIAN

3.1. Metode Penelitian

Penelitian ini menggunakan pendekatan Knowledge Discovery in Databases (KDD) yang

terdiri dari tahapan Data Collection, Data Preprocessing, Data Transformation, Normalisasi Data, Clustering, dan Evaluation. Pendekatan ini digunakan untuk mengekstraksi pola dari data pelanggan untuk menghasilkan segmentasi yang optimal berdasarkan karakteristik transaksi dan profitabilitas [12].

3.2. Data Collection

Data yang digunakan diperoleh dari Kaggle yaitu Indian e-commerce periode 2023-2025 dengan 30.600 data. Dataset terdiri dari beberapa atribut, yaitu *order_id*, *order_date*, *units_sold*, *final_price*, dan *revenue*. Atribut ini digunakan sebagai dasar dalam perhitungan variabel Recency, Frequency, Monetary, dan profitabilitas pelanggan.

3.3. Data Preprocessing

Tahapan preprocessing digunakan untuk meningkatkan kualitas data sebelum dianalisis. Proses meliputi:

a. Data Cleaning

Menghapus data duplikasi, menangani missing value, serta memperbaiki data yang tidak valid.

b. Data Reduction

Memilih atribut yang relevan, yaitu *order_date*, *units_sold*, *final_price*, dan *revenue*.

c. Data Formatting

Menyesuaikan format data pada atribut tanggal agar dapat digunakan untuk menghitung Recency.

3.4. Data Transformation

Pada tahapan ini diperlukan perhitungan variabel utama yang digunakan dalam proses clustering seperti:

a. Recency (R)

Mengukur selisih waktu antara tanggal analisis dengan tanggal transaksi terakhir pelanggan. Semakin kecil nilai R menunjukkan pelanggan semakin aktif.

b. Frequency (F)

Menunjukkan jumlah transaksi yang dilakukan pelanggan dalam periode tertentu.

c. Monetary (M)

Menunjukkan total nilai transaksi yang dilakukan pelanggan.

d. Profitabilitas (P)

Menunjukkan total keuntungan yang diperoleh dari pelanggan. Profitabilitas digunakan untuk melengkapi model RFM sehingga segmentasi lebih mencerminkan kontribusi terhadap perusahaan.

3.5. Normalisasi Data

Data dinormalisasi menggunakan Min-Max Scaling untuk menyamakan skala variabel agar tidak terjadi dominasi atribut tertentu dalam proses clustering.

$$X^t = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3) \quad [9]$$

Digunakan sebagai input pada proses clustering.

3.6. Clustering Dengan K-Means

Proses dilakukan menggunakan algoritma K-Means dengan variabel RFM dan profitabilitas sebagai atribut [13].

Penentuan jumlah cluster menggunakan metode Elbow, yaitu dengan menghitung nilai *Within Clustering Sum of Squares (WCSS)* pada nilai K. Nilai K optimal ditentukan berdasarkan titik dimana penurunan nilai WCSS mulai melambat [9].

Berdasarkan pengujian diperoleh cluster optimal pada K=3, sehingga digunakan 3 cluster.

3.7. Evaluation Clustering

Evaluasi dilakukan menggunakan Silhouette Score untuk mengukur kualitas hasil clustering.

$$S = \frac{b - a}{\max(a, b)} \quad (4) \quad [9]$$

$$S = \frac{b - a}{\max(a, b)}$$

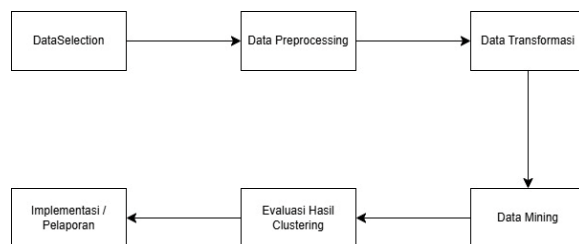
a : Rata-rata data satu cluster
b : Rata-rata jarak ke cluster terdekat
S : Nilai Silhouette Score

3.8. Tools Yang Digunakan

Penelitian menggunakan bahasa pemrograman Python untuk proses pengelolaan data, clustering, serta visualisasi dan menggunakan Google Colab.

3.9. Alur Penelitian

Alur penelitian secara sistematis dimulai dari Data Collection, Data Preprocessing, Data Transformation, Clustering, Evaluation. Alur tersebut digambarkan dalam bentuk flowchart pada Gambar 1.



Gambar 1. Flowchart Alur Penelitian

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Dataset yang digunakan merupakan data e-commerce periode 2023-2025 dengan total 30.600 data dan 15 atribut. Setelah dilakukan seleksi, dihasilkan atribut yang digunakan dalam penelitian meliputi *order_id*, *order_date*, *units_sold*, *final_price*, dan *revenue*.

4.2. Data preprocessing

Tahapan ini dilakukan melalui beberapa proses berikut.

a. Data Cleaning

Pada tahapan ini dilakukan pengecekan terhadap data duplikasi untuk memastikan kualitas data. Hasil menunjukkan bahwa tidak terdapat data duplikasi sehingga dinyatakan bersih.

b. Data Reduction

Tahapan ini dilakukan untuk mengurangi atribut tidak relevan dengan tujuan penelitian.

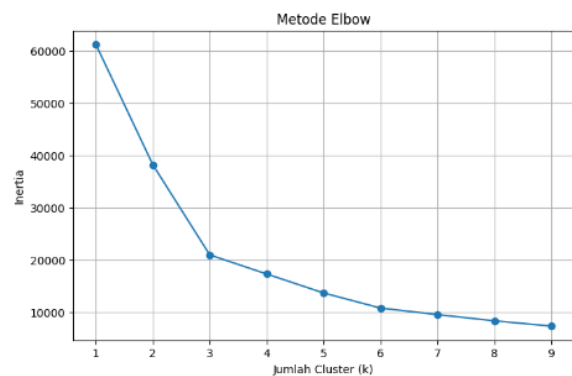
4.3. Data transformation

Data Transformation dilakukan dengan model Recency, Frequency, dan Monetary (RFM) yang bertujuan untuk mengubah data transaksi menjadi bentuk yang lebih terstruktur seperti tabel 4 berikut. Dilanjutkan dengan menormalisasi data untuk menyamakan skala data sehingga setiap variabel memiliki kontribusi yang seimbang.

4.4. Algoritma K-Means clustering

Pada tahapan ini dilakukan pengelompokan data menggunakan algoritma K-Means. Pada clustering diawali dengan normalisasi data dan dilanjutkan dengan penentuan jumlah cluster menggunakan metode Elbow, serta dilakukan perhitungan jarak antar data terhadap centroid menggunakan Euclidean Distance.

Penentuan jumlah cluster dilakukan menggunakan metode Elbow dengan menghitung nilai inertia pada beberapa nilai K. Hasil visualisasi menunjukkan adanya titik siku (*Elbow*) pada nilai K=3. Ditunjukkan pada Gambar 2 visualisasi metode Elbow.



Gambar 2. Visualisasi Metode Elbow

Hasil clustering menghasilkan centroid untuk masing-masing cluster berdasarkan nilai Recency, Frequency, dan Monetary. ditunjukkan pada Tabel 1 berikut.

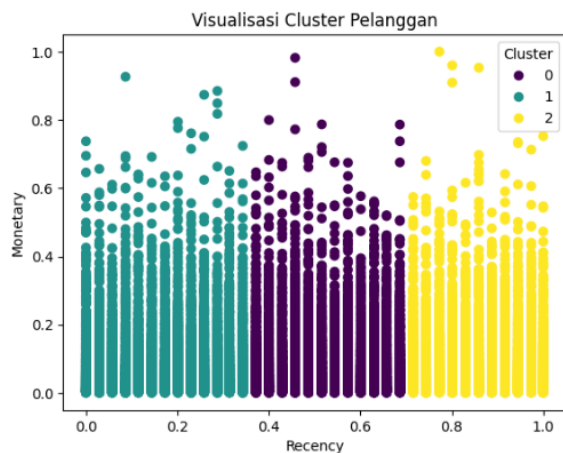
Tabel 1. Penentu Cluster

cluster	Recency	Frequency	Monetary
0	810,48	1,0	48160,72
1	256,17	1,0	48166,05
2	528,20	1,0	263171,34

Penentuan anggota cluster dilakukan dengan menghitung jarak antara data terhadap centroid menggunakan rumus Euclidean Distance.

$$= \sqrt{(R - C_R)^2 + (F - C_F)^2 + (M - C_M)^2} \quad (5) \quad [9]$$

Dihasilkan clustering dengan divisualisasikan untuk melihat distribusi pelanggan berdasarkan nilai Recency, Frequency, dan Monetary. Ditunjukkan pada Gambar 3. Visualisasi Clustering Pelanggan di bawah ini.



Gambar 3. Visualisasi Cluster Customer

Gambar 3 menunjukkan hasil distribusi data pelanggan berdasarkan *Recency*, *Frequency*, dan *Monetary* setelah dilakukan clustering menggunakan algoritma K-Means. Pada visualisasi setiap warna merepresentasikan cluster berdasarkan cluster 0, cluster 1, cluster 2. Penyebaran data menunjukkan bahwa masing-masing cluster memiliki karakteristik berbeda berdasarkan kedekatan nilai terhadap centroid.

Titik centroid merupakan pusat dari masing-masing cluster. Centroid diperoleh dari rata-rata nilai *Recency*, *Frequency*, dan *Monetary* pada setiap cluster. Berdasarkan visualisasi tersebut menunjukkan bahwa:

- Cluster dengan nilai *Monetary* tinggi menunjukkan kelompok pelanggan dengan kontribusi terbesar.
- Cluster dengan nilai *Recency* rendah menunjukkan pelanggan yang lebih aktif atau melakukan transaksi.
- Perbedaan penyebaran antar cluster menunjukkan bahwa metode K-Means berhasil mengelompokkan data berdasarkan kemiripan karakteristik.

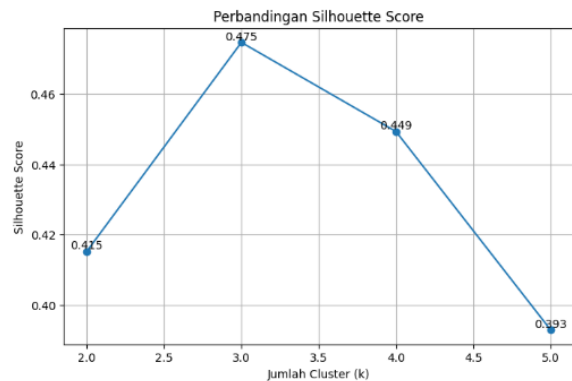
Dengan demikian, visualisasi pada gambar 3 memberikan gambaran jelas mengenai pola distribusi pelanggan pada setiap cluster.

4.5. Evaluasi Data

Pada tahapan ini, evaluasi dilakukan menggunakan metode Silhouette Score yang berguna

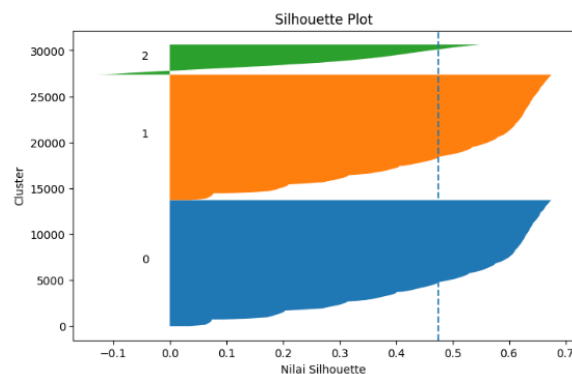
untuk mengukur tingkat kedekatan data dalam satu cluster dan jarak antar cluster.

Perhitungan dilakukan dengan data hasil normalisasi dengan mempertimbangkan cluster dari hasil algoritma K-Means. Nilai Silhouette Score terbaik berada pada K= 3, sehingga jumlah cluster terpilih sebagai jumlah cluster yang optimal. Berikut visualisasi perbandingan Silhouette Score.



Gambar 4. Visualisasi Perbandingan Silhouette Score

Berdasarkan nilai tertinggi jumlah cluster K= 3 diperoleh dengan nilai sebesar 0,4747 dibandingkan dengan jumlah cluster lain. Dengan hasil tersebut menunjukkan bahwa penggunaan 3 cluster memberikan kualitas pengelompokan yang optimal. Berikut Gambar 5 Visualisasi Silhouette.



Gambar 5. Visualisasi Silhouette Plot

Dari gambar di atas menunjukkan bahwa memiliki nilai yang paling mendekati angka 1 dibandingkan dengan jumlah cluster lain. Hal ini menunjukkan bahwa kualitas pengelompokan dengan K=3 merupakan hasil paling optimal dalam merepresentasikan pola data pelanggan.

Berdasarkan hasil tersebut, interpretasi dari masing-masing cluster adalah.

a. Low Loyalty

Terdiri dari pelanggan dengan nilai *Recency* tinggi, *Frequency* rendah dan *Monetary* rendah. Nilai yang tertinggi menunjukkan bahwa pelanggan sudah lama tidak melakukan transaksi, sedangkan nilai *Frequency* dan *Monetary* yang

rendah menunjukkan bahwa pelanggan jarang bertransaksi.

b. Medium Loyalty

Terdiri dari *Recency* sedang, *Frequency* sedang dan *Monetary* sedang. Hal tersebut menunjukkan bahwa pelanggan masih aktif dalam bertransaksi, namun belum berkontribusi sebesar cluster tertinggi. Oleh karena itu dikategorikan pelanggan dengan loyalitas menengah

c. High Loyalty

Terdiri dari pelanggan dengan nilai *Recency* rendah, *Frequency* tinggi, dan *Monetary* tinggi. Menunjukkan bahwa pelanggan sering melakukan transaksi dalam waktu dekat serta memiliki nilai pembelian yang mengindikasikan besar. Menunjukkan potensi loyalitas yang mengindikasikan tinggi dan berpotensi bagi perusahaan.

4.6. Segmentasi Pelanggan

Perbandingan hasil clustering menggunakan algoritma K-Means, dilakukan pengelompokan pelanggan berdasarkan tiga kategori yaitu Low Loyalty, Medium Loyalty, dan High Loyalty. Pengelompokan dilakukan dengan dasar nilai rata-rata (centroid) dari masing-masing cluster. Berikut tabel pengelompokan clusternya.

Tabel 2. Pengelompokan Cluster

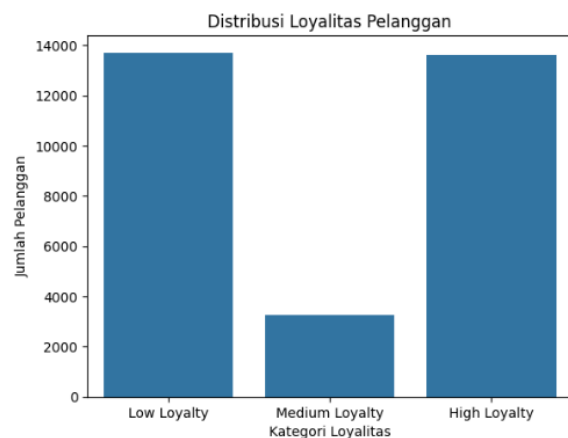
Cluster	R	F	M
0	811,484	1	48160,720
1	257,174	1	48166,052
2	529,203	1	263171,347

Pada tabel 2 di atas dihasilkan pelanggan yang memiliki loyalitas dengan count. Setiap cluster menunjukkan karakteristik perilaku pelanggan yang berbeda, seperti tingkat keseringan transaksi, waktu terakhir bertransaksi, dan total nilai pembelian. Perbedaan nilai tersebut dapat digunakan untuk mengidentifikasi tingkat loyalitas pelanggan, misalnya pelanggan dengan nilai *Monetary* tinggi namun frekuensi rendah atau pelanggan yang lebih aktif bertransaksi. Pada tabel di bawah.

Tabel 3. Persentase Loyalty

Loyalty	Count	Persentase
Low Loyalty	13707	44,8%
Medium Loyalty	13634	44,5%
High Loyalty	3259	10,7%

Pada tabel 3 di atas merupakan tabel Loyalty sehingga dihasilkan distribusi loyalitas pelanggan dengan pengelompokan di bawah ini.



Gambar 6. Distribusi Loyalitas Pelanggan

Berdasarkan gambar 6 menunjukkan sebagian besar pelanggan cenderung dengan tingkat aktivitas rendah dan pelanggan dengan tingkat aktivitas tinggi, sementara pelanggan dengan karakteristik sedang relatif lebih sedikit. Maka dihasilkan strategi pemasaran yang berbeda setiap kategori sebagai berikut.

a. Low Loyalty

Kelompok ini memiliki proporsi jumlah pelanggan yang cukup besar yaitu 44,8 %, sehingga memiliki potensi signifikan untuk meningkatkan total pendapatan apabila diaktivasi kembali. Oleh karena itu, strategi yang diterapkan berfokus pada re-engagement, seperti memberikan diskon ringan sekitar 5%, penawaran khusus, serta kampanye pemasaran berbasis waktu. Strategi ini bertujuan untuk menurunkan nilai *Recency* dan meningkatkan *Frequency*, sehingga pelanggan dapat naik ke kategori Medium Loyalty.

b. Medium Loyalty

Merupakan kontributor terbesar terhadap pendapatan dengan persentase sebesar 44,5%. Hal ini menunjukkan bahwa kelompok ini merupakan segmentasi yang paling potensial untuk dikembangkan. Oleh karena itu, strategi yang diterapkan berfokus pada peningkatan nilai pelanggan, seperti rekomendasi produk berdasarkan riwayat transaksi, pemberian diskon sekitar 10%, serta promo berkala untuk menjaga konsistensi aktivitas pembelian. Strategi ini bertujuan untuk meningkatkan *Frequency* dan *Monetary*, sehingga pelanggan dapat ditingkatkan menjadi kategori High Loyalty.

c. High Loyalty

Meskipun kontribusi revenue kelompok ini mencapai 31%, jumlah pelanggan relatif lebih sedikit dibandingkan kelompok lain. Oleh karena itu, strategi yang diterapkan berfokus pada retensi pelanggan, seperti program membership, pemberian reward berupa gratis ongkir atau penawaran eksklusif, serta strategi upselling untuk meningkatkan nilai transaksi. Strategi ini bertujuan untuk mempertahankan stabilitas

pendapatan serta mencegah pelanggan berpindah ke kompetitor lain.

4.7. Analisis profitabilitas pelanggan

Analisis profitabilitas menghasilkan kontribusi masing-masing cluster terhadap total pendapatan berdasarkan nilai Monetary.

Berdasarkan hasil perhitungan, diperoleh revenue masing-masing cluster sebagai berikut pada tabel 4. Tabel Profitabilitas.

Tabel 4. Tabel Profitabilitas

Cluster	Total Revenue	Persentase	Loyalty
0	720 Juta	33%	Low
1	795 Juta	36%	Medium
2	658 Juta	31%	High

Hasil pada tabel 4. Tabel Profitabilitas menunjukkan bahwa kontribusi revenue antar cluster relatif seimbang. Hasil ini menunjukkan bahwa tingkat Loyalty pelanggan tidak selalu berbanding lurus dengan kontribusi profitabilitas, sehingga diperlukan pendekatan strategi yang berbeda pada setiap segmen. Berdasarkan pola umum, tidak terdapat dominasi dari satu cluster tertentu. Hal ini menunjukkan bahwa distribusi profitabilitas cenderung merata di antara ketiga cluster. Kontribusi pendapatan tidak hanya bergantung pada satu kelompok pelanggan saja, melainkan tersebar pada seluruh segmen.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, penerapan metode *Recency, Frequency, dan Monetary* (RFM) serta algoritma *K-Means Clustering* berhasil mengelompokkan pelanggan ke dalam tiga kategori, yaitu *Low Loyalty*, *Medium Loyalty*, dan *High Loyalty*. Proses penentuan jumlah cluster menggunakan metode *Elbow* menunjukkan bahwa nilai optimal berada pada $K = 3$, yang didukung oleh evaluasi *Silhouette Score* sebesar 0,4747. Hasil tersebut menunjukkan bahwa model *clustering* memiliki kualitas yang baik dalam memisahkan data pelanggan.

Hasil segmentasi menunjukkan bahwa sebagian besar pelanggan berada pada kategori *Low Loyalty* (44,8%), *Medium Loyalty* (44,5%), dan *High Loyalty* (10,7%). Hal ini menunjukkan bahwa mayoritas pelanggan memiliki tingkat aktivitas transaksi yang relatif rendah hingga sedang.

Berdasarkan analisis profitabilitas, diketahui bahwa *revenue* pada setiap cluster relatif seimbang, dengan *Medium Loyalty* sebagai kontributor terbesar (36%), diikuti oleh *Low Loyalty* (33%) dan *High Loyalty* (31%). Temuan ini menunjukkan bahwa tidak terdapat dominasi profit pada satu cluster tertentu, melainkan kontribusi pendapatan tersebar secara merata pada seluruh segmentasi pelanggan.

Dengan demikian, hasil penelitian ini menunjukkan bahwa segmentasi pelanggan tidak hanya mampu mengelompokkan data secara matematis, tetapi juga memberikan gambaran yang

jas mengenai distribusi loyalitas dan profitabilitas pelanggan. Implikasi dari penelitian ini adalah perlunya strategi pemasaran yang tidak hanya berfokus pada satu segmen, tetapi juga mempertimbangkan seluruh kelompok pelanggan secara seimbang guna meningkatkan loyalitas dan kontribusi pelanggan.

DAFTAR PUSTAKA

- [1] F. Arifianto, J. Hasudungan, A. Muzaky, and H. T. Y. Achsan, "Segmentasi Pelanggan Berdasarkan Recency, Frequency, dan Monetary dengan K-Means Clustering: Studi Kasus Toko Pakaian Almost Famous," *Jurnal Teknologi Informatika dan Komputer*, vol. 10, no. 1, pp. 122–135, Mar. 2024, <https://doi.org/10.37012/jtik.v10i1.2096>
- [2] Agung, A., Daniswara, A., Kadek, I., & Nuryana, D. (2023). Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru. *Journal of Informatics and Computer Science*, 05. <https://doi.org/10.26740/jinacs.v5n01.p97-100>
- [3] D. A. Imanuel and G. Alfian, "Visualisasi Segmentasi Pelanggan Berdasarkan Atribut RFM Menggunakan Algoritma K-Means Untuk Memahami Karakteristik Pelanggan pada Toko Retail Online," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 12, no. 2, pp. 283–292, Apr. 2025, doi: 10.25126/jtiik.2025128619.
- [4] T. F. Febriana and M. I. Rosadi, "Analisis Segmentasi Pelanggan Toko Sepatu Di E-Commerce Menggunakan Algoritma K-Means Klastering," *JATI(Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 6, pp. 12069–12072, 2024. <https://doi.org/10.36040/jati.v8i6.11374>
- [5] Ginting, D. E. P., Sitorus Pane, S. A., Nababan, M. N. K., & Christnatis, C. (2025). Smart Diagnosis of Coffee Diseases via Web-Based Expert System. *Sinkron*, 9(3), 1108–1119. <https://doi.org/10.33395/sinkron.v9i3.14974>
- [6] Nugraha, D. S., Thoib, I., Sururi, N., Bayu F, F., & Candra, B. P. (2025). Segmentasi Pelanggan Berbasis Analisis RFM Menggunakan Algoritma K-Means Clustering. *Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI)*, 4(2), 1361–1369. <https://doi.org/10.62712/juktisi.v4i2.641>
- [7] U. Subagyo, A. B. Thoha, A. Syafrianto, H. Santoso, Siswaya, and R. Sanuri, "Customer Behavior Profiling in Wholesale Retail Using RFM Analysis and K-Means Clustering," *Proceedings - 2025 4th International Conference on Electronics Representation and Algorithm: Artificial Intelligence: Creating Tomorrow's World Today, ICERA 2025*, pp. 467–472, 2025, doi: 10.1109/ICERA66156.2025.11087354.
- [8] P. Indra Pangestu, I. Hermanto, D. Irmayanti, S. Tinggi, and T. Wastukencana, "ANALISIS SEGMENTASI PELANGGAN BERBASIS MODEL RECENCY FREQUENCY DAN

- MONETARY (RFM) MENGGUNAKAN ALGORITMA K-MEANS,” *Jl. Cikopak No*, vol. 11, no. 3, p. 214952, Sep. 2023, doi: 10.23960/jitet.v11i3%20s1.3396.
- [9] Y. Syahra, A. Fadlil, and H. Yuliansyah, “Customer Segmentation Using RFM and K-Means Clustering to Support CRM in Retail Industry,” *Sinkron*, vol. 9, no. 3, pp. 1120–1131, 2025, <https://doi.org/10.33395/sinkron.v9i3>
- [10] S. Sajidah, R. Herdiana, and D. Solihudin, “Segmentasi Pelanggan Salon Nuii Beauty Glow Menggunakan K-Means Clustering,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 558–566, 2023, doi: 10.36040/jati. V7i1.6333.
- [11] A. H. Stmik and A. Malang, “Segmentasi Pelanggan Internet Service Provider (ISP) Berbasis Pillar K-Means,” *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 13, no. 2, 2019.