



Analisis Segmentasi Pelanggan Ritel Online Menggunakan K-Means Clustering Berdasarkan Model Recency, Frequency, Monetary (RFM)

Adzriel Yusak Noah Rumapea¹, Dian Pratiwi^{*2}, Syandra Sari³

^{1,2}Informatika, Universitas Trisakti, Jakarta, Indonesia

³Sistem Informasi, Universitas Trisakti, Jakarta, Indonesia

Email Penulis Korespondensi: 2dian.pratiwi@trisakti.ac.id

Abstrak—Segmentasi pelanggan merupakan proses pengelompokan pelanggan berdasarkan karakteristik yang serupa. Segmentasi pelanggan merupakan langkah penting dalam strategi untuk memahami perilaku konsumen dan meningkatkan efektivitas kampanye. Model *Recency, Frequency, Monetary* (RFM) dan algoritma *K-Means Clustering* terbukti dalam mengidentifikasi segmen pelanggan yang berbeda. Penelitian ini menggunakan data transaksi ritel online selama periode 2010-2011. Model RFM digunakan untuk menghitung nilai pelanggan berdasarkan *recency, frequency, monetary*. Selanjutnya, algoritma *K-Means Clustering* diterapkan pada data yang telah dinormalisasi untuk mengelompokkan pelanggan menjadi beberapa segmen. Hasil penelitian didapatkan adanya 3 segmen pelanggan yang berbeda. Ketiga segmen tersebut berbeda-beda secara karakteristik dapat memberikan gambaran yang lebih jelas mengenai perilaku konsumen. Dengan memahami karakteristik setiap segmen, perusahaan dapat merancang strategi pemasaran yang lebih efektif dan personal. Selain itu, penelitian ini juga memberikan kontribusi terhadap pengembangan ilmu pengetahuan dalam bidang analisa data.

Kata Kunci: Segmentasi Pelanggan, Model RFM (*Recency, Frequency, Monetary*), Profil Pelanggan, Algoritma *K-Means Clustering*

Abstract—Customer segmentation is the process of grouping customers based on similar characteristics. It plays a critical role in marketing strategy, enabling businesses to understand consumer behavior and enhance the effectiveness of campaigns. The *Recency, Frequency, Monetary* (RFM) model and the *K-Means Clustering* algorithm have proven to be effective in identifying different customer segments. This study uses online retail transaction data from the period 2010-2011. The RFM model is employed to calculate customer value based on *recency, frequency, and monetary* metrics. Subsequently, the *K-Means Clustering* algorithm is applied to normalized data to group customers into several segments. The results reveal the existence of three distinct customer segments. These segments, characterized by varying traits, provide a clearer understanding of consumer behavior. By understanding the characteristics of each segment, companies can design more effective and personalized marketing strategies. Furthermore, this study contributes to the advancement of knowledge in the field of data analysis.

Keywords: Customer Segmentation, RFM (*Recency, Frequency, Monetary*) Model, Customer Profile, *K-Means Clustering*

I. PENDAHULUAN

Pelanggan kini menjadi penggerak utama bisnis ritel di era digital. Kesuksesan dan profitabilitas bisnis secara langsung terkait dengan pemahaman perilaku pelanggan dan penyesuaian strategi pemasaran untuk memenuhi kebutuhan mereka [1]. Salah satu komponen penting dari pemasaran yang berhasil adalah segmentasi pelanggan, yaitu proses mengelompokkan pelanggan berdasarkan karakteristik yang sama [2]. Dengan melakukan segmentasi pada basis pelanggan, bisnis dapat memperoleh wawasan tentang berbagai segmen konsumen. Hal ini memungkinkan layanan yang lebih individual dan pemasaran yang lebih terarah, yang pada akhirnya meningkatkan kepuasan dan loyalitas pelanggan [3].

Model RFM telah lama digunakan dalam segmentasi pelanggan berdasarkan seberapa baru (*recency*), seberapa sering (*frequency*), dan seberapa besar (*monetary*) interaksi mereka dengan bisnis [3]. Namun, pendekatan ini memiliki keterbatasan dalam menangkap pola yang lebih kompleks dalam data pelanggan. Banyak studi yang menggunakan RFM cenderung menganalisis pelanggan secara sederhana, sehingga hasil segmentasi yang

dihasilkan kurang efektif untuk mendukung strategi pemasaran yang lebih mendalam.

Di sisi lain, metode pengelompokan berbasis seperti *K-Means Clustering* telah terbukti mampu mengklasifikasikan pelanggan secara lebih terperinci dengan memanfaatkan variasi dalam perilaku pelanggan [2][4]. Namun, integrasi antara model RFM dan *K-Means Clustering* untuk segmentasi pelanggan masih belum banyak dieksplorasi dalam konteks ritel daring, terutama dengan data transaksi yang mencakup periode yang panjang. Penelitian ini mengisi kesenjangan tersebut dengan menggabungkan model RFM dan metode *K-Means Clustering* untuk mengidentifikasi segmen pelanggan berdasarkan data transaksi dari sebuah toko ritel daring selama periode Desember 2020 hingga Desember 2021 [3].

Penelitian ini bertujuan untuk mengkaji segmentasi pelanggan ritel daring dengan menggabungkan model RFM dan *K-Means Clustering* untuk mengklasifikasikan pelanggan berdasarkan data transaksi dari sebuah toko ritel daring antara Desember 2020 hingga Desember 2021 [3]. Pertanyaan penelitian utama yang mendorong penyelidikan ini adalah: (1) Segmen pelanggan apa yang dapat diidentifikasi berdasarkan data transaksi? (2) Bagaimana *K-Means Clustering* secara efektif menangkap

Percentage of Missing Values

A horizontal bar chart titled 'Percentage of Missing Values'. The x-axis is labeled 'Percentages (%)' and ranges from 0 to 40 with major grid lines every 5 units. The y-axis lists two categories: 'CustomerID' and 'Description'. The bar for 'CustomerID' is orange and extends to 24.93%. The bar for 'Description' is also orange but is very short, extending to 0.27%.

Variable	Percentage of Missing Values
CustomerID	24.93%
Description	0.27%

2. Penghapusan Data Duplikat

```
[14]: # Displaying the number of duplicate rows
print("The dataset contains {df.duplicated().sum()} duplicate rows that need to be removed.")

# Removing duplicate rows
df.drop_duplicates(inplace=True)

... The dataset contains 5225 duplicate rows that need to be removed.
```

```
[15]: # Getting the number of rows in the dataframe
df.shape[0]

... 401604
```

3. Standardisasi Deskripsi Produk

4. Filterisasi Kolom StockCodes

5. Deteksi dan Penanganan *Outliers*

1. Isolation Forest Algorithm

Algoritma ini umumnya digunakan untuk menemukan data anomali khusus dan data tidak sesuai dengan pola data lainnya dalam sebuah kumpulan data yang besar. Keunggulan algoritma ini adalah kecepatan dalam pemrosesan data dan kemampuan untuk bekerja dengan dataset yang besar. Selain itu, arsitektur paralelnya

[illegible]

Analisis Segmentasi Pelanggan Ritel Online Menggunakan K-Means Clustering Berdasarkan Model Recency, Frequency, Monetary (RFM) **(Adzriel Yusak Noah Rumapea, et all)** 293

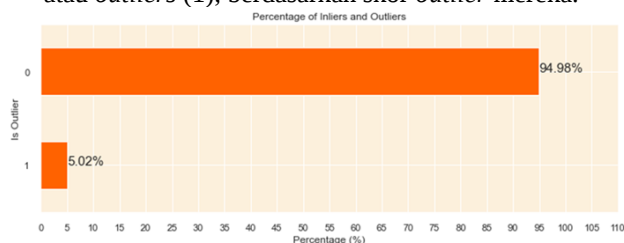


memungkinkan pemanfaatan sumber daya komputasi lebih efektif [7].

Algoritma diinisialisasi dengan parameter kontaminasi sebesar 0,05 yang mengindikasikan ekspektasi 5% data anomali. Nilai ini dipilih untuk menyeimbangkan kualitas data dan menghindari penghapusan data yang berlebihan. Fungsi `fit_predict()` digunakan untuk melatih model pada data yang diberikan dan secara bersamaan menetapkan skor anomali ke setiap titik data. Titik anomali diberi label dengan skor -1, sedangkan titik normal diberi skor 1.

2. Outlier Scoring and Classification

Setelah model terbentuk, skor *outlier* ditambahkan ke dataset dalam kolom baru "Outlier_Scores", yang menunjukkan tingkat deviasi setiap titik dari pola umum. Kolom indikator biner, "Is_Outlier" dibuat untuk mengklasifikasikan titik data sebagai *inliers* (0) atau *outliers* (1), berdasarkan skor *outlier* mereka.



Gambar 4. Persentase *Inliers* dan *Outliers*

3. Visualisasi Proporsi Outlier

Representasi visual berdasarkan kolom "Is_Outlier" menunjukkan bahwa sekitar 5% dari data diklasifikasikan sebagai *outlier*. Proporsi ini dianggap sesuai untuk menyeimbangkan integritas data dan penghapusan titik-titik anomali yang berpotensi mengganggu hasil pengelompokan.

4. Penghapusan Outlier

Untuk meningkatkan kualitas segmentasi, baris-baris yang ditandai sebagai *outlier* dihapus dari dataset. Proses ini memastikan bahwa pengelompokan didasarkan pada pola-pola konsisten yang mewakili basis pelanggan, mengurangi *noise*, dan meningkatkan akurasi kluster.

Rekayasa Fitur Menggunakan Model RFM

Analisis *Recency*, *Frequency*, *Monetary* adalah salah satu metode paling populer yang digunakan dalam segmentasi pelanggan berdasarkan analisis perilaku pelanggan, dan banyak penelitian telah mengimplementasikan model RFM dalam segmentasi pelanggan. Model RFM pertama kali diusulkan oleh Arthur Hughes pada tahun 1994, dan seiring berjalannya waktu penggunaan model ini menjadi semakin lazim [8].

Untuk memfasilitasi segmentasi pelanggan yang bermakna, penelitian ini menggunakan model *Recency*, *Frequency*, *Monetary* (RFM), sebuah alat yang banyak digunakan dalam analisis perilaku pelanggan. Model RFM memungkinkan evaluasi nilai pelanggan berdasarkan aktivitas pembelian mereka, memungkinkan bisnis untuk mengkategorikan pelanggan berdasarkan tingkat keterlibatan dan potensi pendapatan mereka. Dalam penelitian ini, model RFM memberikan metrik penting

untuk pengelompokan, memanfaatkan tiga komponen utama dari data transaksi.

1. Recency (R)

Recency mengukur lamanya waktu sejak transaksi terakhir pelanggan. Nilai yang lebih tinggi mengindikasikan interval waktu yang lebih lama sejak transaksi terakhir, sementara nilai yang lebih rendah menunjukkan pembelian yang lebih baru. Jumlah hari antara tanggal pembelian terbaru setiap pelanggan dengan tanggal akhir analisis digunakan untuk menghitung nilai *recency*. Karena pembeli baru-baru ini menunjukkan minat aktif dan memiliki peluang lebih tinggi untuk melakukan lebih banyak pembelian di masa depan, indikator ini penting dalam menentukan tingkat keterlibatan konsumen.

2. Frequency (F)

Frekuensi mengukur seberapa sering seorang pelanggan melakukan pembelian dalam jangka waktu tertentu. Angka ini ditentukan dengan menjumlahkan semua transaksi yang dilakukan oleh setiap pelanggan, yang memungkinkan identifikasi pelanggan yang sering melakukan pembelian. Karena mereka menunjukkan interaksi yang sering dengan bisnis, skor lebih tinggi menunjukkan loyalitas konsumen.

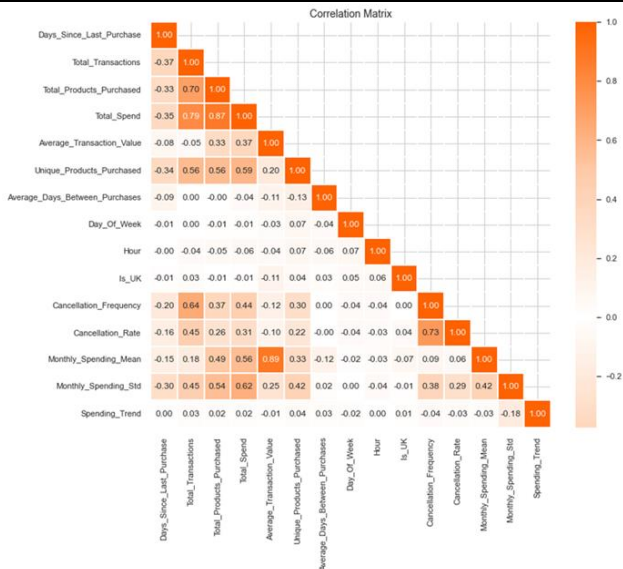
3. Monetary (M)

Skor moneter mencerminkan total pengeluaran yang dilakukan oleh setiap konsumen dalam periode waktu tertentu. Angka ini diperoleh dengan menjumlahkan total pendapatan dari pembelian setiap pelanggan. Pelanggan dengan nilai moneter yang lebih tinggi dianggap sebagai pelanggan bernilai tinggi dan memberikan kontribusi signifikan terhadap total pendapatan.

Analisis Korelasi

Analisis korelasi digunakan untuk memeriksa kekuatan dan arah hubungan linear antara variabel-variabel terkait pelanggan dalam dataset "customer_data". Teknik statistik ini menilai tingkat asosiasi antara variabel, yang direpresentasikan oleh koefisien korelasi yang berkisar dari -1 hingga 1. Koefisien korelasi positif mengindikasikan hubungan langsung, di mana kedua variabel meningkat atau menurun bersamaan. Sebaliknya, koefisien negatif menunjukkan hubungan terbalik, di mana satu variabel meningkat sementara yang lain menurun. Koefisien yang lebih dekat ke 1 (positif atau negatif) menunjukkan korelasi yang lebih kuat, sedangkan koefisien yang lebih dekat ke nol mengindikasikan hubungan linear yang lebih lemah atau tidak ada.

Sebuah *heatmap* dibuat untuk merepresentasikan secara visual hubungan antara variabel-variabel. *Heatmap* menampilkan koefisien korelasi dan kekuatannya yang sesuai menggunakan gradien warna. Warna yang lebih terang menunjukkan korelasi yang lebih kuat, sementara warna yang lebih gelap menunjukkan hubungan yang lebih lemah. Sel-sel putih mewakili variabel yang tidak memiliki korelasi.

Gambar 5. *Heatmap Correlation*

Algoritma Klusterisasi : K-Means Clustering

Kemampuan pengelompokan untuk membagi data menjadi kelompok-kelompok diskrit berdasarkan kesamaan fitur menjadi alasan pemilihannya. Untuk meminimalkan jumlah kuadrat dalam-cluster, algoritma secara iteratif menetapkan titik data ke *centroid* kluster terdekat sebelum menghitung ulang *centroid*. Tindakan-tindakan berikut diambil:

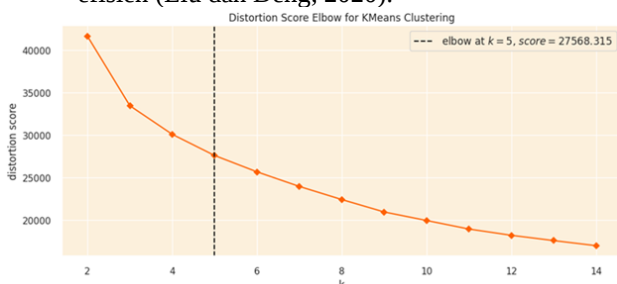
1. Menentukan Nilai K Optimal

Menentukan jumlah kluster sangat penting untuk segmentasi yang efektif. Untuk menentukannya, diimplementasikan dua metode berikut :

a) Metode Siku / *Elbow Method*

Metode ini merupakan metodologi untuk menentukan jumlah kluster yang seharusnya digunakan berdasarkan proporsi nilai perbandingan antar jumlah kluster. Metode ini menghitung WCSS (*Within Cluster Sum of Squares*) untuk setiap hasil kluster (Muningsih & Kiswati, 2018).

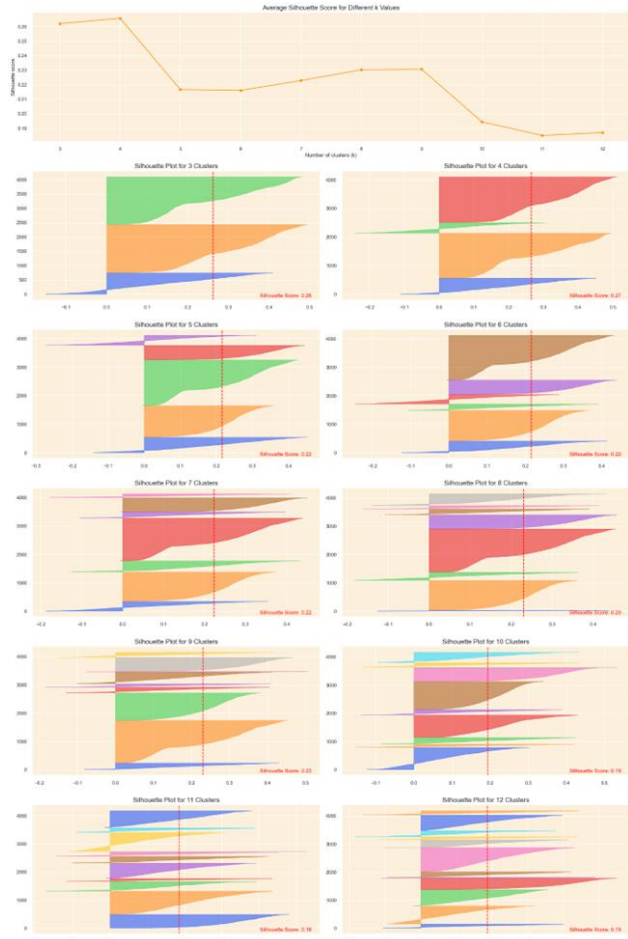
WCSS mengidentifikasi jumlah kluster yang optimal dalam *K-Means Clustering*. Prinsipnya adalah dengan mencari titik penurunan WCSS secara drastis. Titik ini menandai keseimbangan antara kedekatan data dengan *centroid* dan kompleksitas model. Dengan demikian, metode *elbow* (Aminazadeh dan Noorbebhahani, 2022) menjadi alat yang efektif dalam penentuan parameter K dan mencapai segmentasi data yang efisien (Liu dan Deng, 2020).

Gambar 6. *Elbow Method Graphic*

Grafik tersebut mengusulkan 5 sebagai jumlah cluster optimal (K). Namun, "titik siku" yang khas yang biasanya terlihat pada grafik semacam ini tidak terlihat jelas dalam kasus ini. Nilai inersia, yang mengukur dispersi dalam-cluster, terus menurun hingga K=5, menunjukkan bahwa K optimal mungkin terletak antara 3 dan 7. Untuk menyempurnakan perkiraan ini, analisis siluet, metode evaluasi komplementer, dapat diterapkan.

b) Metode Siluet / *Silhouette Method*

Skor siluet dihitung untuk setiap nilai k untuk mengevaluasi pemisahan antara cluster. Nilai koefisien siluet rata-rata yang tinggi menunjukkan pengelompokan yang baik, membantu dalam pemilihan nilai k optimal [12]. Dapat disimpulkan bahwa jumlah cluster optimal yang dapat dibentuk adalah K = 3. Karena cluster pada K=3 sudah di atas nilai rata-rata.



Gambar 7. Grafik Siluet

2. Formasi Kluster

Setelah seluruh proses pengelompokan selesai, kolom baru berlabel 'cluster' akan dimasukkan ke dalam DataFrame data pelanggan yang telah dibersihkan. Kolom yang baru ditambahkan ini akan berisi nomor cluster yang ditetapkan (0, 1, atau 2) untuk setiap titik data, secara efektif memetakan setiap titik data ke cluster yang sesuai.



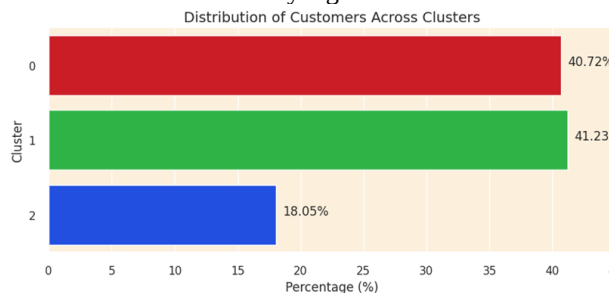
Python	
Spending_Trend	cluster
0.0	1
4.486071	2
-100.884	0
0.0	1
0.0	1

Gambar 8. Hasil Pengelompokkan

III. HASIL DAN PEMBAHASAN

Visualisasi Distribusi Jumlah Kluster

Visualisasi distribusi kluster merupakan teknik untuk memvisualisasikan data bagaimana data di distribusikan di antara kluster yang berbeda.



Gambar 9. Grafik Persentase Distribusi Kluster

Distribusi pelanggan di masing-masing kluster menunjukkan proporsi yang signifikan. Dari hasil pengelompokkan menggunakan algoritma K-Means Clustering, tiga kluster utama terbentuk :

1. Kluster 0
Kluster ini mencakup 40,72% pelanggan, menunjukkan pelanggan dengan frekuensi pembelian yang moderat tetapi nilai transaksi yang rendah.
2. Kluster 1
Kluster ini mencakup 41,23% pelanggan, dengan nilai transaksi tinggi tetapi frekuensi pembelian lebih rendah.
3. Kluster 2

Kluster ini mencakup 18,05% pelanggan, dengan frekuensi pembelian tinggi dan nilai transaksi yang signifikan.

Matriks Evaluasi

1. Silhouette Score

Merupakan ukuran jarak di antara kluster. Semakin tinggi nilainya maka semakin baik separasi kluster yang ada. Nilai ini berkisar dari -1 sampai dengan 1. Skor mendekati +1 menunjukkan kelompok terpisah dengan baik. Skor 0 menunjukkan bahwa antar klasternya tidak signifikan atau kluster yang tumpang tindih. Dan skor -1 menunjukkan bahwa kluster tersebut kemungkinan besar tidak terbentuk dengan baik.

Hasil pengukuran didapatkan skor 0.2362 menunjukkan kluster yang lebih berbeda dan terpisah dengan baik dari kluster lainnya

2. Calinski Harabasz Score

Metrik ini berguna untuk melakukan evaluasi terhadap penyebaran di dalam dan di luar kluster. Semakin tinggi skor yang didapatkan maka semakin baik pendefinisian kluster tersebut [13].

Pengukuran metrik ini didapatkan 1257,17 yang menunjukkan kluster terdefinisi dengan baik dan menunjukkan pengelompokkan yang terjadi berhasil menemukan struktur substansial dalam data.

3. Davies Bouldin Score

Metrik ini menilai rata-rata kesamaan satu kluster dengan kluster lainnya yang paling mirip.

Skor yang didapatkan untuk metrik Davies Bouldin sebesar 1.37 menunjukkan tingkat kesamaan yang moderat antara setiap kluster dengan kluster yang paling miripnya. Skor lebih rendah berarti lebih sedikit kesamaan antar kluster.

Analisis Kluster dan Profil

Pada bagian akan dilakukan analisis terhadap karakteristik yang ada pada setiap kluster. Hal ini dilakukan guna untuk mengetahui dengan lebih mendalam perilaku dan preferensi yang berbeda-beda. Selain itu, dibuat *profiling* dari setiap kluster untuk mengidentifikasi sifat-sifat penting pelanggan di dalam tiap kluster.

1. Radar Chart Approach

Digunakan untuk memvisualisasikan nilai *centroid* dari setiap kluster. Grafik berikut ini memberikan perbandingan visual dengan cepat tentang profil dari berbagai kluster.



Gambar 10. Radar Chart Setiap Kluster

- Kluster 0 (Red Chart)

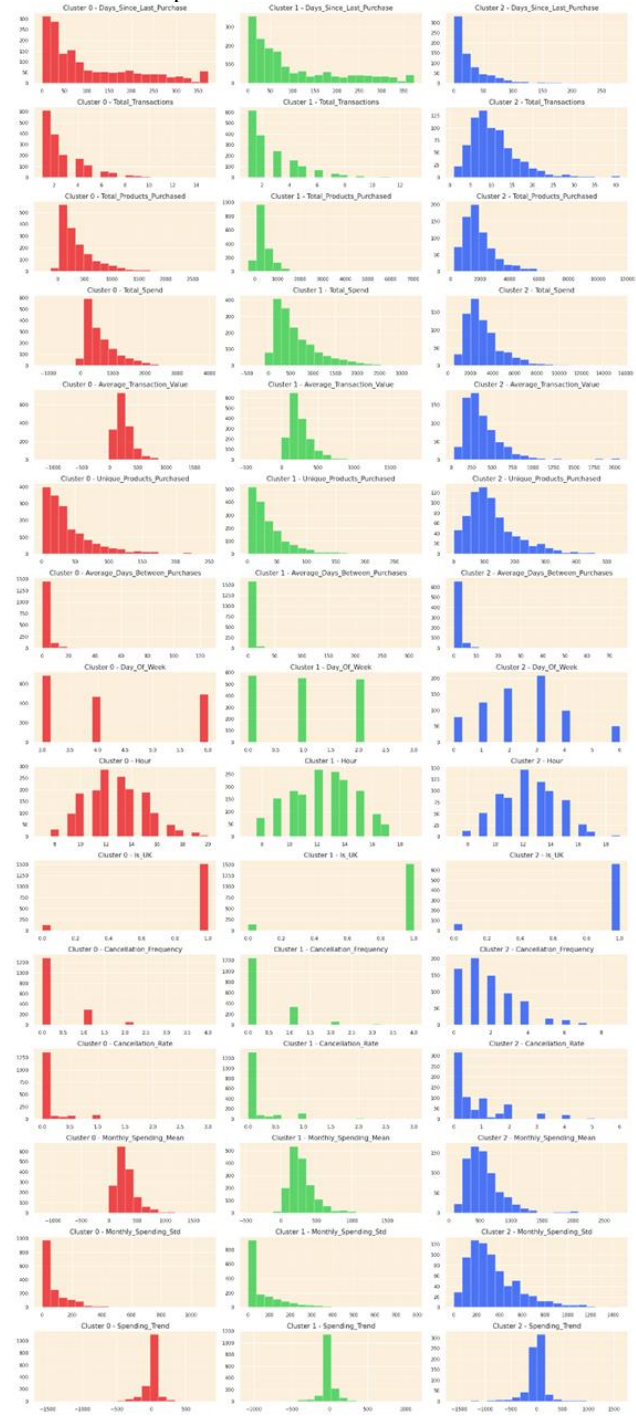
Profil pelanggan untuk kluster 0 terdiri dari pelanggan yang melakukan pembelian



sesekali, dengan preferensi untuk berbelanja pada akhir pekan. Profil lengkap pelanggan dalam klaster ini adalah sebagai berikut :

- Pelanggan dalam klaster ini cenderung memiliki jumlah transaksi yang rendah dan membeli produk dalam jumlah kecil.
 - Perilaku pembelian mereka menunjukkan kecenderungan yang lebih rendah untuk berbelanja pada akhir pekan, seperti terlihat dari nilai pada variabel *Day_of_Week* yang tinggi.
 - Pelanggan dalam klaster ini jarang membatalkan pesanan, seperti yang ditunjukkan oleh tingkat pembatalan yang rendah.
 - Nilai transaksi rata-rata mereka relatif rendah, menunjukkan bahwa mereka cenderung menghabiskan lebih sedikit dalam setiap pembelian.
- **Klaster 1 (Green Chart)**
Profil pelanggan dalam cluster 1 ditandai dengan pembelian besar yang jarang terjadi. Berikut adalah profil lengkap pelanggan dalam cluster ini:
- Pelanggan dalam cluster ini menunjukkan pengeluaran moderat, ditambah dengan frekuensi transaksi yang lebih rendah. Hal ini terlihat dari nilai yang tinggi pada *Days_Since_Last_Purchase* dan *Average_Days_Between_Purchases*.
 - Meskipun transaksi jarang terjadi, ada tren kenaikan pengeluaran pelanggan dari waktu ke waktu.
 - Pelanggan dalam cluster ini lebih suka berbelanja pada sore atau malam hari, seperti yang ditunjukkan oleh nilai *Hour* yang tinggi, dan sebagian besar dari mereka tinggal di Inggris Raya.
- **Klaster 2 (Blue Chart)**
Pelanggan dalam cluster 2 menunjukkan pengeluaran tinggi dan frekuensi pembelian tinggi, tetapi juga tingkat pembatalan yang tinggi. Berikut adalah profil lengkap pelanggan dalam cluster ini:
- Pelanggan dalam cluster ini adalah pembelanja besar dengan variasi produk yang tinggi.
 - Pelanggan melakukan transaksi sering tetapi juga memiliki tingkat pembatalan pesanan yang tinggi.
 - Pelanggan ini memiliki waktu rata-rata yang singkat antara pembelian dan cenderung berbelanja di pagi hari, seperti yang ditunjukkan oleh nilai *Hour* yang rendah.
2. Histogram Chart Approach
Untuk memvalidasi profil yang diidentifikasi oleh *radar chart*, dapat dilakukan membuat *plot histogram* untuk setiap

fitur yang disegmentasi berdasarkan label klaster. Dengan menggunakan histogram memungkinkan untuk mengidentifikasi secara visual distribusi nilai fitur dalam setiap klaster, sehingga hasil yang didapatkan dari *radar chart* dapat disempurnakan.



Gambar 11. Histogram Seluruh Fitur

Perbandingan Dengan Penelitian Terdahulu

Hasil penelitian ini dibandingkan dengan studi sebelumnya tentang segmentasi pelanggan yang menggunakan *K-Means Clustering*. Perbandingan ini



mengungkap kesamaan dan perbedaan yang signifikan dengan temuan penelitian lain.

1. Harani et al. (2020) mengidentifikasi tiga kelompok pelanggan. Salah satu kelompok yang diidentifikasi adalah pelanggan loyal dengan frekuensi pembelian tinggi dan tingkat pembatalan rendah, memiliki kemiripan dengan klaster 2 dalam penelitian ini. Meskipun demikian, terdapat perbedaan mendasar dalam konteks produk yang diteliti. Peneliti berfokus kepada perilaku konsumen dalam konteks ritel *online*. Dan Harani dkk. berfokus kepada produk *digital*.
2. Hoang Thi Ha dan Pradana (2021) mengidentifikasi lima segmen pengunjung mal dengan pola pengeluaran yang berbeda-beda. Hasil penelitian tersebut menunjukkan bahwa kelompok pelanggan dengan pendapatan dan pengeluaran tinggi merupakan target yang paling potensial untuk kampanye pemasaran eksklusif. Pelanggan dalam segmen ini cenderung lebih responsif terhadap promosi yang dipersonalisasi dan layanan tambahan. Temuan ini sejalan dengan hasil penelitian kami, di mana Klaster 1, yang terdiri dari pelanggan dengan pembelian besar namun frekuensi pembelian rendah, dapat menjadi target utama untuk penawaran produk eksklusif atau program bundling.
3. Romadansyah et al. (2021) Analisis segmentasi pelanggan *e-commerce* menggunakan model LRFM mengungkapkan bahwa pelanggan dengan frekuensi pembelian tinggi dan nilai transaksi besar (*core customer*) memiliki kontribusi signifikan terhadap peningkatan pendapatan perusahaan dalam jangka panjang. Penelitian ini menyimpulkan bahwa mengidentifikasi pelanggan aktif dan berpotensi tinggi (sebanding dengan Klaster 2 dalam penelitian ini) merupakan langkah krusial dalam merancang strategi pemasaran yang efektif, khususnya melalui program loyalitas dan personalisasi.

IV. KESIMPULAN

Penelitian ini menunjukkan efektivitas penggabungan model Recency, Frequency, Monetary, dan algoritma K-Means Clustering dalam segmentasi pelanggan ritel daring. Dengan menggunakan data transaksi dari Desember 2010 hingga Desember 2011, tiga klaster pelanggan berbeda yang berbeda berhasil diidentifikasi. Distribusi ini mencerminkan bahwa metode yang digunakan mampu menghasilkan pelanggan berdasarkan pola perilakunya. Ditinjau dari parameter metrik evaluasi yang digunakan, menunjukkan bahwa klaster yang terbentuk terpisah secara jelas. Temuan utama penelitian ini meliputi :

1. Efektivitas penggabungan model RFM dan algoritma K-Means Clustering

Integrasi keduanya dapat menghasilkan wawasan yang mendalam mengenai perilaku pelanggan, seperti mengidentifikasi pelanggan bernilai tinggi dan merancang strategi pemasaran yang terarah.

2. Strategi pemasaran lebih personal terhadap setiap segmen

Setiap segmen menunjukkan karakteristik unik yang dapat menjadi dasar strategi pemasaran. Dapat dibuat strategi program loyalitas untuk Klaster 2 dan promosi yang lebih intensif untuk meningkatkan frekuensi pembelian di Klaster 1.

3. Relevansi data jangka panjang

Penggunaan data dalam durasi satu tahun penuh memungkinkan analisis yang lebih representatif terhadap perilaku pelanggan dibandingkan dengan studi sebelumnya yang hanya menggunakan data dengan rentang waktu yang singkat.

Penelitian ini menyoroti kekuatan segmentasi berbasis data dalam ritel *online*. Dengan menerapkan model RFM-K-Means, peneliti berhasil mengungkap wawasan berharga tentang perilaku dan preferensi pelanggan. Wawasan ini dapat dimanfaatkan untuk mengembangkan kampanye pemasaran yang ditargetkan, rekomendasi produk yang dipersonalisasi, dan strategi retensi pelanggan yang efektif.

Setiap segmen pelanggan yang diidentifikasi menunjukkan perilaku unik yang dapat menginformasikan strategi pemasaran yang ditargetkan. Misalnya, pelanggan bernilai tinggi dapat diberi penghargaan dengan program loyalitas, sementara pelanggan yang tidak aktif dapat dilibatkan kembali melalui kampanye yang dipersonalisasi.

Penelitian ini berkontribusi pada bidang analisis data dengan menunjukkan efektivitas algoritma pengelompokan dalam segmentasi pelanggan ritel *online*. Penggunaan Metode Elbow dan skor siluet memastikan keandalan hasil segmentasi, memungkinkan retailer untuk membuat keputusan berbasis data

Dengan mengintegrasikan sumber data tambahan dan menerapkan teknik pembelajaran mesin yang lebih canggih, penelitian di masa depan dapat lebih menyempurnakan dan meningkatkan model segmentasi pelanggan yang disajikan dalam penelitian ini. Hal ini dapat menghasilkan wawasan pelanggan yang lebih tepat dan strategi pemasaran yang lebih baik.

Kesimpulannya, penelitian ini menunjukkan efektivitas pendekatan RFM-K-Means dalam segmentasi pelanggan *online*. Dengan memanfaatkan teknik ini, retailer dapat memperoleh wawasan berharga tentang perilaku pelanggan, menyesuaikan strategi pemasaran, dan meningkatkan kepuasan pelanggan.

UCAPAN TERIMA KASIH

Penulis ingin mengucapkan terima kasih kepada Dosen Pembimbing yang telah memberikan bimbingan, petunjuk, dukungan moril, dan saran yang berharga selama proses penelitian dan penulisan artikel ini. Selain itu, tidak



lupa juga kepada seluruh dosen dan tenaga pendidik di Program Studi Informatika Universitas Trisakti yang telah memberikan pengetahuan dan ilmu selama menempuh pendidikan. Pengalaman dan pembelajaran yang diberikan menjadi landasan utama penulis menyelesaikan artikel ini.

Penulis juga mengucapkan terima kasih kepada keluarga untuk doa, semangat, dukungan moril dan materil. Keluarga menjadi kekuatan bagi penulis dalam menghadapi hambatan dan kesulitan yang muncul selama proses penulisan.

Ada banyak pihak yang terlibat dalam penulisan skripsi ini, yang tidak dapat disebutkan satu per satu. Penulis menyadari bahwa bantuan mereka, baik secara langsung maupun tidak langsung sangat penting untuk penyelesaian penelitian ini.

V. REFERENSI

- [1] Abadi, S., et al. "Application model of k-means clustering: Insights into promotion strategy of vocational high school." *International Journal of Engineering & Technology*, vol. 7, no. 2.27 Special Issue, 2018, pp. 182–187. doi: 10.14419/ijet.v7i2.11491.
- [2] Chaudhary, P. "Data Mining System, Functionalities and Applications: A Radical Review." *International Journal of Distributed and Parallel Systems*, vol. 1, no. 1, 2015, pp. 32–44.
- [3] Harani, N. H., Prianto, C., and Nugraha, F. A. "Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python." *Jurnal Manajemen Informatika*, vol. 10, no. 2, 2020, pp. 133–146. doi: 10.34010/jamika.v10i2.2683.
- [4] Puspitasari, N., Widiars, J. A., and Setiawan, N. B. "Customer segmentation using bisecting k-means algorithm based on recency, frequency, and monetary (RFM) model." *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 2, 2020, pp. 78–83. doi: 10.14710/jtsiskom.8.2.2020.78-83.
- [5] Siagian, R., Sirait, P. S. P., and Halima, A. "E-Commerce Customer Segmentation Using K-Means Algorithm and Length, Recency, Frequency, Monetary Model." *Journal of Informatics Telecommunication Engineering*, vol. 5, no. 1, 2021, pp. 21–30. doi: 10.31289/jite.v5i1.5182.
- [6] R. D. Hastuti, I. Waspada, P. W. Wirawan, and N. Bahtiar, "Designing computer-assisted problem-based learning (CAPBL) environment for performance analysis of isolation forest algorithm," *J. Phys. Conf. Ser.*, vol. 1524, no. 1, 2020, doi: 10.1088/1742-6596/1524/1/012103.
- [7] H. Ge and L. Jin, "Human Posture Recognition Based on Isolation Forest Algorithm," *Adv. Transdiscipl. Eng.*, vol. 37, pp. 276–284, 2023, doi: 10.3233/ATDE230149.
- [8] G. ASLANTAŞ, M. GENÇGÜL, M. RUMELLI, M. ÖZSARAÇ, and G. BAKIRLI, "Customer Segmentation Using K-Means Clustering Algorithm and RFM Model," *Deu Muhendis. Fak. Fen ve Muhendis.*, vol. 25, no. 74, pp. 491–503, 2023, doi: 10.21205/deufmd.2023257418.
- [9] Muningsih, E., & Kiswati, S. (2018). Sistem aplikasi berbasis optimasi metode elbow untuk penentuan clustering pelanggan. *Joutica*, 3(1), 117.
- [10] Aminazadeh, M., & Noorbehbahani, F. (2022). City intersection clustering and analysis based on traffic time series. In 2022 12th International Conference on Computer and Knowledge Engineering (ICCKE) (pp. 274-281). IEEE. <https://doi.org/10.1109/ICCKE57176.2022.9960065>
- [11] Liu, F., & Deng, Y. (2020). Determine the number of unknown targets in open world based on elbow method. *IEEE Transactions on Fuzzy Systems*, 29(5), 986-995.
- [12] A. Z. Faridee and V. P. Janeja, "Measuring Peer Mentoring Effectiveness," *Am. J. o.*, vol. 15, no. 2, pp. 7–22, 2020.
- [13] N. H. Harani, C. Prianto, and F. A. Nugraha, "Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python," *J. Manaj. Inform.*, vol. 10, no. 2, pp. 133–146, 2020, doi: 10.34010/jamika.v10i2.2683.
- [14] Pradana, M. G., & Ha, H. T. (2021). *Maximizing Strategy Improvement in Mall Customer Segmentation using K-means Clustering*. *Journal of Applied Data Sciences*, 2(1), 19–25. <https://doi.org/10.47738/jads.v2i1.18>
- [15] R. Siagian, P. S. Pahala Sirait, and A. Halima, "E-Commerce Customer Segmentation Using K-Means Algorithm and Length, Recency, Frequency, Monetary Model," *J. Informatics Telecommun. Eng.*, vol. 5, no. 1, pp. 21–30, 2021, doi: 10.31289/jite.v5i1.5182.
- [16] J. V. Hutagalung, A. Sediyo, and B. Solihah, "Analysis of Topic Movement & Conversation Membership on Twitter using K-Means Clustering," vol. 4, no. 2, pp. 67–73, 2024.