

VISUALISASI SEGMENTASI PELANGGAN BERDASARKAN ATRIBUT RFM MENGUNAKAN ALGORITMA K-MEANS UNTUK MEMAHAMI KARAKTERISTIK PELANGGAN PADA TOKO RETAIL ONLINE

Dennis Alfa Imanuel¹, Ganjar Alfian*²

^{1,2} Departemen Teknik Elektro dan Informatika, Sekolah Vokasi, Universitas Gadjah Mada, Yogyakarta

Email: ¹dennisalfa01@mail.ugm.ac.id, ²ganjar.alfian@ugm.ac.id

*Penulis Korespondensi

(Naskah masuk: 2 Desember 2024, diterima untuk diterbitkan: 11 April 2025)

Abstrak

Meningkatnya minat belanja pelanggan toko *retail online* menimbulkan persaingan ketat antar *retailer*. Agar tetap unggul dan kompetitif, *retailer* perlu memahami karakteristik pelanggannya. Penerapan segmentasi pelanggan memberikan kemudahan pada *retailer* untuk memahami karakteristik pelanggan berdasarkan penilaian pada atribut yang dihitung dari data riwayat transaksi pelanggan. Hasil segmentasi pelanggan yang divisualisasikan dapat meningkatkan pemahaman *retailer* dalam memahami data dan membantu dalam proses pengambilan keputusan. Oleh karena itu, penelitian ini mengusulkan Visualisasi Segmentasi Pelanggan menggunakan Algoritma *K-means* berdasarkan Atribut RFM (*Recency, Frequency, Monetary*). Hasil segmentasi dapat digunakan untuk Memahami Karakteristik Pelanggan pada Toko *Retail Online*. Penelitian ini menggunakan algoritma *k-means* untuk menjalankan *clustering* yang performanya akan dibandingkan dengan algoritma *k-medoids* mengacu pada nilai *silhouette*, *Calinski-Harabasz Index*, dan *DaviesBouldin Index* dalam melakukan segmentasi pelanggan berdasarkan atribut RFM. Berdasarkan metrik tersebut, didapatkan nilai algoritma *k-means* berturut-turut adalah 0,6558, 0,7219, dan 3578,9, sedangkan nilai algoritma *k-medoids* adalah 0,4677, 0,8298, dan 1236,9. Dengan demikian, hasilnya menunjukkan bahwa kinerja *clustering* menggunakan *k-means* lebih baik daripada menggunakan *k-medoids*. Pada *dashboard Looker Studio* ditampilkan visualisasi data hasil segmentasi tersebut, kemudian diuji fungsionalitasnya dengan metode *Blackbox Testing* dan berhasil menyelesaikan semua skenario pengujian, kemudian dilakukan pengujian dengan metode UAT (*User Acceptance Testing*) dan mendapatkan predikat sangat layak.

Kata kunci: visualisasi data, looker studio, segmentasi pelanggan, *k-means*, RFM, *retail online*

VISUALIZATION OF CUSTOMER SEGMENTATION BASED ON RFM ATTRIBUTES USING K-MEANS ALGORITHM TO COMPREHEND CUSTOMER CHARACTERISTICS WITHIN AN ONLINE RETAIL STORE

Abstract

The growing interest in online retail shopping among customers has resulted in intense competition among retailers. To sustain a competitiveness, retailers need to understand characteristics of their customer. Implementation of customer segmentation facilitates retailers in understanding customer characteristics through assessments based on attributes derived from customer transaction history data. Visualization of customer segmentation results can enhance the retailer's understanding of data and assist in the decision-making process. Therefore, this study proposes the Visualization of Customer Segmentation using the K-means Algorithm based on RFM Attributes (*Recency, Frequency, Monetary*). The segmentation results can be utilized to understand the characteristics of customers in an online retail store. This study explores the *k-means* algorithm to execute clustering, and its performance will be compared with the *k-medoids* algorithm, based of *silhouette* values, *Calinski-Harabasz Index*, and *Davies Bouldin Index* in customer segmentation based on RFM attributes. Based on given metrics, the consecutive performance values for *k-means* algorithm are 0.6558, 0.7219, and 3578.9, while *k-medoids* algorithm are 0.4677, 0.8298, and 1236.9. Thus, the results indicate that the clustering performance using *k-means* is better than using *k-medoids*. On the *Looker Studio* dashboard, the visualization of the segmentation data is displayed, and its functionality is tested using the *Black Box Testing* method, successfully completing all test scenarios. Subsequently, the system undergoes testing through the *User Acceptance Testing* (UAT) method and receives a highly satisfactory rating.

Keywords: data visualization, looker studio, customer segmentation, *k-means*, RFM, online retail

1. PENDAHULUAN

Transaksi jual beli eceran atau *retailing* adalah kegiatan transaksi yang melibatkan penjual yang dalam konteks ini adalah *retailer* dan pembeli yang dalam konteks ini adalah pelanggan (Severin et al., 2001). *Retailing* merupakan tahapan akhir dalam rantai pasokan, karena pada tahap ini produk dan jasa dipindahkan dari tangan produsen atau pemasok ke tangan pelanggan akhir (Dunne et al., 2013). Kegiatan *retailing* sudah tertanam erat dalam kehidupan hidup bermasyarakat dan memiliki peran penting pada perekonomian dengan menyediakan kemudahan akses bagi pelanggan untuk memenuhi kebutuhan sehari-hari akan produk dan jasa (Bhatia, 2008).

Berkembangnya teknologi internet dan perangkat digital telah membuka peluang baru bagi *retailer* bersamaan dengan peningkatan minat belanja pelanggan *retail online* (Barile et al., 2018). Peningkatan tersebut menimbulkan persaingan antar *retailer* menjadi semakin ketat dan intensif (Zaharia & Schmitz, 2020). Dengan demikian, diperlukan identifikasi karakteristik dan cara untuk mempertahankan pelanggan agar dapat terus bersaing (Mutandwa et al., 2009).

Kegiatan *retailing online* memiliki keunikan, yaitu proses, aktivitas, minat, dan preferensi berbelanja pelanggan saat membeli produk atau menikmati layanan jasa dapat dilacak dan dikumpulkan (Jai et al., 2013). Dengan data tersebut, *retailer* dapat pendekatan bisnis secara personal dan beragam berdasarkan karakteristik pelanggan dengan melakukan segmentasi (D. Chen et al., 2012). Segmentasi pelanggan membantu *retailer* untuk mengelompokkan karakteristik dari tiap-tiap individu pelanggan pada sejumlah *cluster* berdasarkan kesamaan pada variabel maupun atribut tertentu (Chang et al., 2010).

Segmentasi pelanggan berdasarkan atribut RFM (*recency*, *frequency*, dan *monetary*) dapat menggambarkan karakteristik pelanggan berdasarkan data riwayat transaksi pelanggan (Miglautsch, 2000). Pengelompokan pelanggan ke dalam segmen dapat dilakukan dengan menerapkan metode *clustering* dalam *unsupervised learning* menggunakan algoritma *k-means* karena kecepatan, efisiensi, dan kemudahan penerapannya (A. H. L. Chen et al., 2022). Untuk meningkatkan pemahaman *retailer* akan hasil segmentasi pelanggan menggunakan algoritma *k-means* berdasarkan atribut RFM diperlukan *dashboard* visualisasi data. Dengan adanya *dashboard* visualisasi data, *retailer* dapat mengidentifikasi pola, memahami informasi yang terkandung, dan menarik sebuah keputusan (Sadiku et al., 2016).

Pada penelitian sebelumnya, segmentasi pelanggan berdasarkan atribut RFM diimplementasikan menggunakan beberapa metode, seperti menerapkan algoritma *k-means* dan mencari jumlah optimal *cluster* berdasarkan nilai *silhouette*

(Anitha & Patil, 2022). Penelitian lain juga melakukan segmentasi pelanggan menggunakan algoritma *k-means*, akan tetapi mencari jumlah optimal *cluster* berdasarkan *elbow method* (A. H. L. Chen et al., 2022). Penelitian tersebut menerapkan algoritma *k-means* untuk melakukan segmentasi pelanggan dengan jumlah optimal *cluster* yang ditentukan berdasarkan metrik tertentu, akan tetapi belum ada metrik yang menilai performa algoritma *k-means* yang diterapkan.

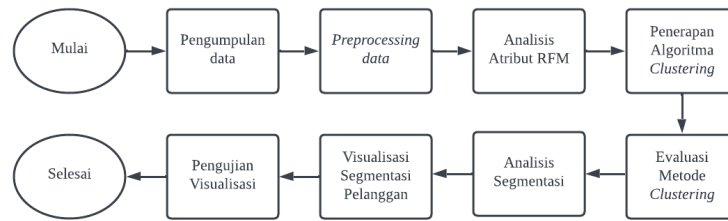
Penelitian ini akan melakukan segmentasi pelanggan berdasarkan atribut RFM menggunakan algoritma *k-means* dan membuat *dashboard* visualisasi untuk menampilkan hasilnya. Performa algoritma *k-means* akan dibandingkan dengan algoritma *k-medoids* dalam menjalankan segmentasi akan diuji berdasarkan tiga *cluster validation metrics*, yaitu nilai *silhouette*, *Calinski-Harabasz Index*, dan *DaviesBouldin Index*. Keunikan penelitian ini terletak pada integrasi model clustering dengan *dashboard* visualisasi, yang performanya dievaluasi melalui pengujian fungsional menggunakan metode blackbox testing serta pengujian kelayakan melalui user acceptance testing (UAT).

2. METODE PENELITIAN

Dalam menjalankan penelitian ini, terdapat delapan tahapan utama pengerjaannya yang ditampilkan menggunakan *flowchart* pada Gambar 1.

Tahap pertama, yaitu pengumpulan data dilakukan dengan dengan mempelajari *dataset* yang digunakan untuk penelitian, dengan nama *Online Retail II* dari *UC Irvine Machine Learning Repository* yang mendokumentasikan transaksi secara *online* untuk bisnis *retail* tanpa toko di United Kingdom (D. Chen, 2019). *Dataset* berbentuk *spreadsheet* ini memuat daftar transaksi antara 01 Desember 2009 hingga 09 Desember 2011, akan tetapi penelitian ini hanya menggunakan data transaksi antara 01 Desember 2010 hingga 09 Desember 2011 yang dimuat pada sheet bernama Year 2010-2011. Dengan memilih hanya menggunakan data pada sheet tersebut, memungkinkan analisis yang lebih mendalam pada rentang tersebut. Adapun faktor pendukung lainnya adalah untuk mempercepat runtime pemrograman dan menjalankan *dashboard* visualisasi data dengan lebih cepat karena beban komputasi yang lebih ringan, sehingga penelitian dapat dijalankan dengan lebih efektif, menghindari error, dan meningkatkan pengalaman pengguna.

Setelah data telah dikumpulkan, diperlukan tahapan kedua, yaitu *preprocessing* data. Tahapan ini mengolah dan meningkatkan kualitas data mentah untuk menjaga konsistensi dan ketepatan hasil analisis data (Fan et al., 2021).



Gambar 1. Diagram alir penelitian

Tahapan ini dilakukan dengan cara menghapus atribut *dataset* yang tidak akan digunakan, menghapus baris dengan nilai kosong, mengubah tipe data, dan menambahkan atribut yang diperlukan pada penelitian.

Data yang telah diolah akan di analisis atribut RFM-nya. Tahapan ketiga ini dilakukan dengan mengidentifikasi atribut dari *dataset* yang memiliki karakteristik sesuai dengan kebutuhan analisis. Atribut yang memenuhi karakteristik yang diinginkan akan diekstraksi untuk menghitung nilai *recency*, *frequency*, dan *monetary* (Sharyanto & Lestari, 2022). Atribut dengan karakteristik RFM tersebut akan diekstraksi kepada *dataframe* baru berdasarkan data dari masing-masing pelanggan dan diubah penamaan atributnya sehingga sesuai dengan atribut RFM untuk mempermudah proses penelitian. Untuk menjelaskan proses tersebut, tahapan ini akan diterapkan dengan contoh data riwayat transaksi pelanggan yang ditunjukkan pada Tabel 1.

Tabel 1. Contoh *Dataset*

CustomerID	PurchaseDate	TotalAmount
1001	01/01/01	\$ 10
1002	02/02/01	\$ 20
1003	03/01/01	\$ 10
1004	04/01/01	\$ 5
1001	02/02/01	\$ 10
1004	05/01/01	\$ 5

Berdasarkan atribut pada contoh *dataset* tersebut, atribut *frequency* didapatkan dengan menghitung jumlah transaksi atribut CustomerID, atribut *monetary* didapatkan dengan menghitung jumlah TotalAmount untuk setiap CustomerID, dan atribut *recency* didapatkan dengan persamaan 1.

$$recency = \frac{1}{a + 1} \quad (1)$$

Di mana, a merupakan selisih tanggal pada kolom PurchaseDate terakhir setiap CustomerID dengan tanggal pada kolom PurchaseDate terakhir dalam *dataset* (dalam hitungan bulan). Dengan demikian, didapatkan nilai RFM untuk tiap CustomerID, yang ditunjukkan pada Tabel 2.

Tabel 2. Contoh Perhitungan RFM

CustomerID	Recency	Frequency	Monetary
1001	1	2	20
1002	1	1	20
1003	0,5	1	10
1004	0,5	2	10

Dataframe dengan nilai pada kolomnya memuat atribut RFM tersebut, akan dinormalisasikan menggunakan *z-score*. *Z-score* digunakan dengan tujuan untuk membuat persebarannya lebih relevan sehingga meningkatkan kualitas *cluster* yang akan dibuat dalam proses analisis data (Sarmah et al., 2011). Dengan menggunakan *z-score*, nilai-nilai data akan diubah sehingga memiliki rata-rata nol dan standar deviasi satu, seperti yang ditampilkan pada Tabel 3.

Tabel 3. Contoh Normalisasi Data

CustomerID	Recency	Frequency	Monetary
1001	1	1	1
1002	1	-1	1
1003	-1	-1	-1
1004	-1	1	-1

Tahapan keempat, yaitu penerapan algoritma *clustering* dilakukan dengan membuat segmentasi pelanggan berdasarkan algoritma *k-means* dan *k-medoids*. Agar algoritma tersebut dapat dijalankan, penelitian ini menerapkan *elbow method* untuk menentukan jumlah optimal *cluster* dari algoritma *k-means* dan nilai *silhouette* untuk menentukan jumlah optimal *cluster* dari algoritma *k-medoids*.

Elbow method merupakan metode untuk menggambarkan presentase variasi yang dihasilkan terhadap jumlah *cluster* (k) dalam sebuah fungsi. Fungsi akan menambahkan jumlah k hingga menemukan nilai optimal (Syakur et al., 2018). Nilai k akan ditambahkan satu persatu dalam fungsi dan nilai *sum square error* (SSE) akan dihitung. SSE merupakan jumlah dari rata-rata *euclidean distance* dari data point (x) ke *centroid* pada *cluster* ke- k (C_k), seperti yang ditunjukkan pada persamaan 2 (Marutho et al., 2018).

$$SSE = \sum_{K=1}^K \sum_{xi \in Sk} \|xi - Ck\|_2^2 \quad (2)$$

Nilai *silhouette* (*s*) pada *data point* (*i*) dihitung dengan menentukan perbedaan *intra-cluster* (*a*) dan *inter-cluster* (*b*) dari setiap sampel, kemudian mendefinisikan *silhouette coefficient* dari setiap sampel berdasarkan perbedaan ini, seperti yang ditunjukkan pada persamaan 3.

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

$$s(i) = \begin{cases} 1 - \frac{a(i)}{b(i)}, & a(i) < b(i) \\ 0, & a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1, & a(i) > b(i) \end{cases} \quad (3)$$

Setelah jumlah optimal *cluster* didapatkan, maka penerapan algoritma *clustering* sudah mungkin untuk dilakukan. Langkah perhitungan dalam menjalankan algoritma *k-means* dimulai dengan memilih *centroid* (*y*) awal secara acak dan menghitung jaraknya dengan *data point* (*x*) menggunakan *euclidean distance* (*d*), seperti yang ditunjukkan pada persamaan 4 (Puspitasari et al., 2022).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

Langkah berikutnya adalah mengelompokkan *data point* ke dalam *cluster* berdasarkan jarak terkecil dengan *centroid*. Langkah ketiga adalah menghitung ulang jarak dengan nilai *centroid* baru berbeda yang didapatkan dari nilai rata - rata *data point* pada masing-masing *cluster*. Setelahnya, mengulangi perhitungan dari langkah pertama hingga *centroid* tetap dan anggota data pada *cluster* tidak berubah (Faturahim & Baita, 2022).

Penerapan algoritma *k-medoids* memiliki kemiripan dengan *k-means*, perbedaannya terletak pada *centroid*-nya terletak pada *data point* dan disebut sebagai *medoids* (Wang et al., 2013). Langkah pertama adalah memilih *data point* secara acak sebagai *medoids* awal (*y*) dan menghitung jaraknya dengan *data point* (*x*) menggunakan persamaan *manhattan distance* (*d*), seperti yang ditunjukkan pada persamaan 5.

$$d(x, y) = (\sum_{i=1}^n |x_i - y_i|) \quad (5)$$

Langkah kedua, mengelompokkan *data point* ke dalam *cluster* berdasarkan jarak terdekat dengan *medoids* dan menghitung *total cost*. *Total cost* adalah jumlah jarak semua *data point* dengan *medoids* pada *cluster* yang sama. Langkah ketiga, menghitung jarak kandidat *medoids* dengan *data point* dan menghitung *total cost*. Kandidat *medoids* merupakan *data point* lainnya yang memiliki jarak terdekat pada masing-masing *cluster* yang belum dijadikan *medoid*. Langkah keempat, menghitung total simpangan (*S*) dengan menghitung selisih *total cost* pada masing-

masing *cluster* kandidat *medoids* dengan *cluster medoids*. Apabila *S* < 0, ganti *medoids* dengan kandidat *medoids*, jika tidak tetap gunakan *medoids awal*. Kemudian, melakukan perulangan perhitungan dari langkah ketiga hingga *medoids* tetap dan anggota data pada *cluster* tidak berubah.

Tahapan kelima adalah evaluasi metode *clustering*. Evaluasi metode *clustering* bertujuan untuk membandingkan performa *means* algoritma *k-means* dengan algoritma *k-medoids* berdasarkan tiga *cluster validation metrics*, yaitu nilai *silhouette*, *Calinski-Harabasz Index*, dan *DaviesBouldin Index*. *Calinski-Harabasz index* (*CH_k*) merupakan metrik validitas yang membandingkan perhitungan persebaran data antar *cluster* (*BCSM*) dan dalam *cluster* (*WCSM*). Semakin tinggi nilai dari *CH_k*, maka *cluster* semakin optimal, seperti yang ditunjukkan pada persamaan 6 (Martino & Rossetto, 2022).

$$CH_k = \frac{BCSM}{k-1} \times \frac{n-k}{WCSM} \quad (6)$$

Davies bouldin index (*DBI*) adalah metrik validasi yang digunakan untuk mengevaluasi kinerja algoritma *clustering*. Indeks ini mengukur kesamaan antara suatu *cluster* (*i*) dengan *cluster* lainnya (*j*) yang paling mirip dengan mempertimbangkan nilai maksimum dari rasio antar *cluster* (*maxR(i, j)*). Nilai *DBI* yang lebih rendah menunjukkan *clustering* yang lebih baik, seperti yang ditunjukkan pada persamaan 7 (Ashari et al., 2022).

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{j \neq i} R(i, j) \quad (7)$$

Tahapan keenam adalah analisis segmentasi, yang dilakukan dengan menganalisis karakteristik pelanggan berdasarkan tinggi atau rendahnya nilai rata-rata atribut RFM-nya pada segmentasi pelanggan yang dihasilkan oleh algoritma *k-means* yang telah dinilai performanya. Penelitian ini akan melakukan penelusuran literatur terkait tipe-tipe pelanggan yang dihasilkan oleh masing-masing *cluster* untuk dilakukan peninjauan karakteristiknya.

Tahapan ketujuh, yaitu visualisasi segmentasi pelanggan. Pada tahap ini, hasil analisis segmentasi akan ditampilkan pada *dashboard looker studio*. Halaman *dashboard* pada *looker studio* dapat dibuat dan dibagikan untuk memantau *key performance indicators*, mengidentifikasi pola data, dan mendapatkan wawasan untuk pengambilan keputusan dan tindakan (Saadeh et al., 2022). Visualisasi data pada *dashboard* di *platform looker studio* terdiri dari empat halaman, yaitu halaman utama, halaman analisis pelanggan, halaman analisis segmentasi, dan halaman *top chart*.

Tahapan kedelapan atau terakhir adalah pengujian visualisasi. Pengujian dilakukan untuk menilai performa *dashboard* visualisasi dengan melalui metode *blackbox testing* dan UAT. *Blackbox testing* juga disebut sebagai *functional testing*

prosedur untuk pengujian berdasarkan informasi dari spesifikasi dan fungsionalitas perangkat lunak (Nidhra & Dondeti, 2012). *User acceptance testing* (UAT) merupakan prosedur untuk pengujian dari solusi yang diberikan sistem perangkat lunak agar bersesuaian dengan kebutuhan pengguna dengan mengumpulkan umpan balik dan mengidentifikasi masalah di dalamnya sebelum dipublikasikan (Al-Hurmuzi et al., 2018). UAT melibatkan pengguna dan pengembang yang sebelumnya telah diberikan tutorial tentang penggunaan sistem tersebut untuk menentukan kelayakan perangkat lunak dengan memberikan survei berupa kuisioner (Arif & Soko, 2022).

3. HASIL DAN PEMBAHASAN

3.1. Pengujian Performa Algoritma

Pengujian performa algoritma dalam melakukan *clustering* pada *dataset* ini dinilai berdasarkan perbandingan nilai *cluster validation metrics*, dengan hasil pengujian ditunjukkan pada Tabel 4.

Tabel 4. Performa *Clustering*

<i>Cluster Validation Metrics</i>	<i>K-Means</i>	<i>K-Medoids</i>
Nilai <i>Silhouette</i>	0,6558	0,4677
<i>DaviesBouldin Index</i>	0,7219	0,8298
<i>Calinski-Harabasz Index</i>	3578,9	1236,9

Hasil pengujian tersebut menunjukkan bahwa performa algoritma *k-means* lebih baik dibandingkan algoritma *k-medoids* pada penerapannya di penelitian ini, ditunjukkan dari nilai *silhouette* dan *Calinski-Harabasz Index* yang lebih tinggi, akan tetapi *DaviesBouldin Index* lebih rendah.

Dalam *k-means*, *centroid* dihitung sebagai rata-rata titik data dalam *cluster*, sehingga lebih sensitif terhadap distribusi data dan mampu meminimalkan nilai *intra-cluster* lebih efektif. Dalam penelitian ini, pelanggan memiliki pola distribusi data yang relatif terpusat dan homogen, sehingga pendekatan *k-means* lebih mampu mengidentifikasi struktur *cluster* yang optimal dibandingkan dengan *k-medoids*, tercermin pada nilai *silhouette* dan *Calinski-Harabasz Index* yang lebih tinggi, serta *Davies-Bouldin Index* yang lebih rendah.

3.2. Hasil Analisis Segmentasi

Penerapan *elbow method* saat melakukan *clustering* dengan algoritma *k-means* menghasilkan jumlah optimal *cluster* tiga. Nilai rata-rata dari setiap atribut RFM pada ketiga *cluster* ditampilkan untuk menunjukkan karakteristik pelanggan pada masing-masing *cluster* ditunjukkan pada Tabel 5.

Cluster 0, memiliki rata-rata nilai *recency*, *frequency*, dan *monetary* rendah menunjukkan pelanggan pada *cluster* ini sudah lama tidak

berbelanja, jarang berbelanja, dan memiliki nilai transaksi yang sedikit. Virgiawan (2015),

Tabel 5. Hasil Segmentasi

<i>Cluster</i>	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>
0	0,29	2,43	961
1	0,98	66,5	85.904
2	0,998	6,23	2.481

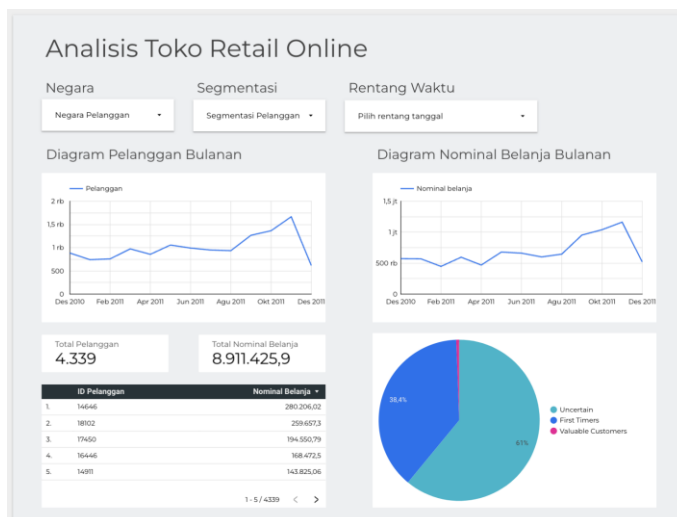
mengategorikan pelanggan tersebut dalam segmentasi *uncertain*. Segmentasi *uncertain* merujuk kepada karakteristik pelanggan yang memiliki tingkat ketidakpastian dalam perilaku dan keputusan pembelian mereka. Pelanggan ini cenderung sulit diprediksi dan mempengaruhi tingkat permintaan yang tidak stabil. Segmentasi ini cenderung menjadi kelompok terbesar, seringkali mewakili setengah dari basis pelanggan namun hanya menyumbang sedikit pendapatan (Marcus, 1998).

Cluster 1, memiliki rata-rata nilai *recency*, *frequency*, dan *monetary* tinggi menunjukkan pelanggan pada *cluster* ini terakhir berbelanja dalam jangka waktu yang dekat, sering berbelanja, dan memiliki nilai transaksi yang banyak. Virgiawan (2015), mengategorikan pelanggan tersebut dalam segmentasi *valuable*. Segmentasi *valuable* merujuk kepada karakteristik pelanggan yang memiliki nilai penting bagi bisnis karena potensinya dalam menghasilkan pendapatan dan kemampuan untuk memberikan *feedback* yang bisnis. Banyak perusahaan mengadopsi *customer relationship management* (CRM) untuk bertahan dalam lingkungan yang kompetitif dengan mempertahankan dan mengembangkan pelanggan pada segmentasi ini (Morisada et al., 2019).

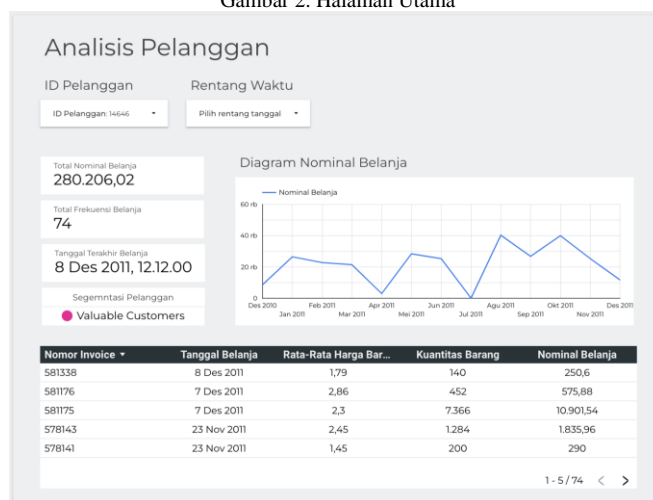
Cluster 2, memiliki rata-rata nilai *recency* tinggi, tetapi nilai *frequency* dan *monetary* rendah menunjukkan pelanggan pada *cluster* ini terakhir berbelanja dalam jangka waktu yang dekat, akan tetapi tidak sering berbelanja sehingga nilai transaksinya sedikit. Virgiawan (2015), mengategorikan pelanggan tersebut dalam segmentasi *first timers*. Segmentasi *first timers* merujuk kepada karakteristik pelanggan yang baru atau pertama kali melakukan pembelian atau menggunakan jasa dari suatu bisnis. Segmentasi ini cenderung memperhatikan atribut khusus dan kelebihan yang ditawarkan pada produk (Parasuraman, 1997).

3.3. Hasil Visualisasi Data

Hasil analisis segmentasi pelanggan akan divisualisasikan menggunakan *dashboard looker studio*. *Dashboard* memuat empat halaman yang menampilkan analisis dari data riwayat transaksi pelanggan, yaitu halaman halaman utama, analisis pelanggan, analisis segmentasi, dan top chart.



Gambar 2. Halaman Utama



Gambar 3. Halaman Analisis Pelanggan

Halaman utama atau halaman analisis toko menampilkan riwayat kinerja toko dalam diagram pelanggan dan nominal belanja untuk bulanan, menampilkan total pelanggan dan nominal belanja secara keseluruhan, tabel berisikan urutan pelanggan dengan nominal belanja tertinggi, dan diagram proporsi segmentasi dengan tiga *filter* data berdasarkan negara pelanggan, segmentasi pelanggan, dan rentang waktu transaksi. Tampilan halaman ini ditunjukkan pada Gambar 2. Halaman analisis pelanggan pada *dashboard* menampilkan analisis mendalam yang terpusat pada seorang pelanggan. Halaman ini bertujuan untuk mencari data setiap individu pelanggan dengan memasukkan Customer ID pada *filter* dan menampilkan total nominal, frekuensi, dan tanggal terakhir belanja pelanggan (atribut RFM), diagram nominal belanja, dan riwayat transaksi untuk setiap transaksi yang dilakukan oleh pelanggan. Halaman ini secara *default* akan menampilkan data dari pelanggan dengan id pelanggan 14646. Tampilan halaman ini ditunjukkan pada Gambar 3.

Halaman analisis segmentasi pada *dashboard* menampilkan analisis atribut RFM mendalam yang terpusat pada masing-masing hasil analisis

segmentasi. Halaman ini bertujuan untuk menampilkan persebaran negara pelanggan pada setiap segmen dan penjelasan mendalam terkait atribut RFM melalui tampilan diagram atribut RFM, diagram geografis, dan menampilkan rata-rata nilai *monetary*, *frequency*, dan *recency* dilengkapi *filter* data berdasarkan segmentasi dan rentang waktu belanja. Tampilan halaman ini ditunjukkan pada Gambar 4. Halaman Top Chart menampilkan empat tabel berisikan peringkat yang mengurutkan id pelanggan dengan nilai terbaik untuk masing-masing atribut RFM dan nomor *invoice* dengan nominal belanja terbanyak. Halaman ini bertujuan untuk memberikan kemudahan akses untuk menemukan profil-profil pelanggan yang memiliki nilai terbaik pada masing-masing atribut RFM dan mengidentifikasi transaksi dengan nominal belanja terbanyak. Halaman ini memiliki tiga *filter* data berdasarkan negara pelanggan, segmentasi pelanggan, dan rentang waktu transaksi. Tampilan halaman ini ditunjukkan pada Gambar 5.

3.4. Pengujian Dashboard Visualisasi

Pada halaman *dashboard* *looker studio* yang digunakan untuk menampilkan hasil analisis tersebut,

dua metode pengujian, yaitu *blackbox testing* dan UAT. Pada penerapan metode pengujian *blackbox testing* diberikan delapan skenario pengujian dengan *test case* mengakses halaman maupun menerapkan *filter* data untuk menilai fungsionalitas dari *dashboard*, seperti yang ditunjukkan pada Tabel 6.

Tabel 6. *BlackBox Testing*

No	Halaman Dashboard	Skenario Pengujian	Test Case	Hasil
1	Halaman utama	Mengkases halaman	Mengakses halaman utama dengan klik <i>link dashboard</i>	Berhasil menampilkan data
2		Menerapkan <i>filter</i> data	Menerapkan <i>filter</i> berdasarkan negara, segmentasi, dan rentang waktu transaksi	Berhasil mengubah data
3	Halaman analisis pelanggan	Mengkases halaman	Mengakses halaman analisis pelanggan melalui panel	Berhasil menampilkan data
4		Menerapkan <i>filter</i> data	Menerapkan <i>filter</i> berdasarkan id pelanggan dan rentang waktu transaksi	Berhasil mengubah data
5	Halaman analisis segmentasi	Mengkases halaman	Mengakses halaman analisis pelanggan melalui panel	Berhasil menampilkan data
6		Menerapkan <i>filter</i> data	Menerapkan <i>filter</i> berdasarkan segmentasi dan rentang waktu transaksi	Berhasil mengubah data
7	Halaman <i>top chart</i>	Mengkases halaman	Mengakses halaman analisis	Berhasil menampilkan data

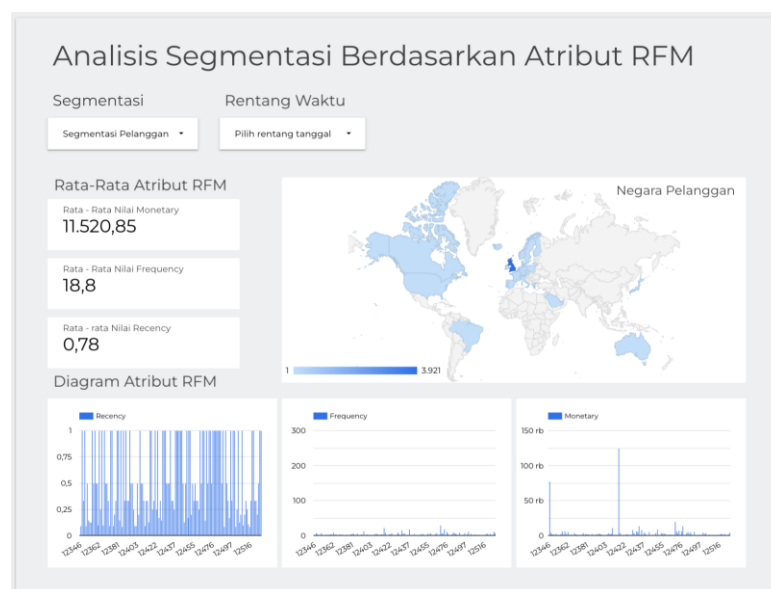
8	Menerapkan <i>filter</i> data	pelanggan melalui panel Menerapkan <i>filter</i> berdasarkan negara, segmentasi, dan rentang waktu transaksi	Berhasil mengubah data
---	-------------------------------	--	------------------------

4. KESIMPULAN DAN SARAN

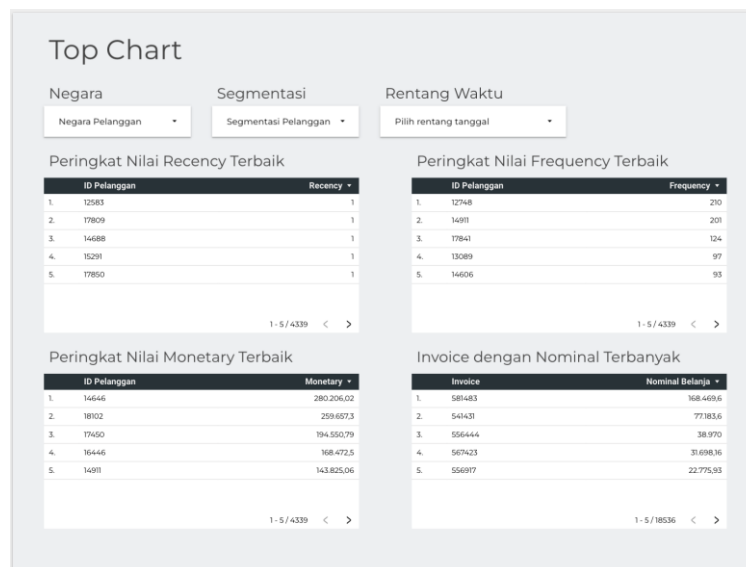
Penelitian ini melakukan segmentasi pelanggan pada *dataset* riwayat transaksi pada toko *retail online* dan menemukan bahwa penerapan algoritma *k-means* memiliki performa yang lebih baik apabila dibandingkan dengan penerapan algoritma *k-medoids*, dengan jumlah segmentasi yang ideal sebanyak tiga. Segmentasi *uncertain* menunjukkan pelanggan dengan karakteristik nilai RFM rendah, *valuable* menunjukkan pelanggan dengan karakteristik nilai RFM tinggi, dan *first timers* menunjukkan pelanggan dengan karakteristik nilai *recency* tinggi, tapi *frequency* dan *monetary* rendah.

Dashboard visualisasi data yang menampilkan hasil segmentasi dapat menjalankan semua skenario pengujian fungsionalitas dan memiliki predikat sangat layak.

Pengembangan untuk penelitian selanjutnya sangat memungkinkan untuk menambahkan analisis metode *preprocessing* data, pengujian fitur keamanan, dan eksplorasi metode pemasaran pada masing-masing karakteristik pelanggan.



Gambar 4. Halaman Analisis Segmentasi

Gambar 5. Halaman *Top Chart***DAFTAR PUSTAKA**

- AL-HURMUZI, S., AL-KHANJARI, Z., & AI-KINDI, I. 2018. Proposed Feasible PEF framework for User Acceptance Testing. 2018 8th International Conference on Computer Science and Information Technology (CSIT), 242–248.
<https://doi.org/10.1109/CSIT.2018.8486225>
- ANITHA, P., & PATIL, M. M. 2022. RFM model for customer purchase behavior using K-Means algorithm. *Journal of King Saud University - Computer and Information Sciences*, 34(5), 1785–1792.
<https://doi.org/https://doi.org/10.1016/j.jksuci.2019.12.011>
- ARIF, E., & SOKO, I. P. 2022. The Evaluation of web-based and android face-to-face tutorial applications quality using the user acceptance testing (UAT) method. *Journal of World Science*.
<https://api.semanticscholar.org/CorpusID:251992255>
- ASHARI, I. F., BANJARNAHOR, R., FARIDA, D. R., AISYAH, S. P., DEWI, A. P., & HUMAYA, N. 2022. Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies. *Journal of Applied Informatics and Computing*.
<https://api.semanticscholar.org/CorpusID:250371540>
- BARILE, S., POLESE, F., & SARNO, D. 2018. Grocery retailing in the I4. 0 era. *Symphonya. Emerging Issues in Management*, 2, 38–51.
- BHATIA, S. C. 2008. *Retail management*. Atlantic Publishers & Dist.
- CHANG, E.-C., HUANG, S.-C., & WU, H.-H. 2010. Using K-means method and spectral clustering technique in an outfitter's value analysis. *Quality & Quantity*, 44, 807–815.
- CHEN, A. H. L., LIANG, Y.-C., CHANG, W.-J., SIAUW, H.-Y., & MINANDA, V. 2022. RFM model and K-means clustering analysis of transit traveller profiles: a case study. *Journal of Advanced Transportation*, 2022.
- CHEN, D. 2019. Online Retail II. UCI Machine Learning Repository.
- CHEN, D., SAIN, S. L., & GUO, K. 2012. Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*, 19, 197–208.
- DUNNE, P. M., LUSCH, R. F., & CARVER, J. R. 2013. *Retailing*. Cengage Learning.
- FAN, C., CHEN, M., WANG, X., WANG, J., & HUANG, B. 2021. A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data. *Frontiers in Energy Research*.
<https://api.semanticscholar.org/CorpusID:232387214>
- FATUROHIM, I., & BAITA, A. 2022. Penerapan Algoritma Kmeans Clustering Untuk Mengelompokkan Kerawanan Rob Di Daerah Pekalongan. *JuTI "Jurnal Teknologi Informasi"*.
<https://api.semanticscholar.org/CorpusID:251793749>
- ISTIYAN, N., NYOTO, R. D., & MUHARDI, H. 2020. Aplikasi Learning Management System Pada Jenjang Madrasah Aliyah. *Justin (Jurnal Sistem Dan Teknologi Informasi)*, 8(1), 105–115.
- JAI, T.-M. C., BURNS, L. D., & KING, N. J. 2013. The effect of behavioral tracking practices on consumers' shopping evaluations and repurchase intention toward trusted online

- retailers. *Computers in Human Behavior*, 29(3), 901–909.
- MARCUS, C. 1998. A practical yet meaningful approach to customer segmentation. *Journal of Consumer Marketing*, 15(5), 494–504. <https://doi.org/10.1108/07363769810235974>
- MARTINO, A., & ROSSETTO, A. 2022. A Hybrid Score to Optimize Clustering Hyperparameters for Online Search Term Data. 2022 IEEE International Conference on Big Data (Big Data), 2317–2322. <https://doi.org/10.1109/BigData55660.2022.10020638>
- MARUTHO, D., HANDAKA, S. H., WIJAYA, E., & MULJONO. 2018. The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News. 2018 International Seminar on Application for Technology of Information and Communication, 533–538. <https://doi.org/10.1109/ISEMANTIC.2018.8549751>
- MIGLAUTSCH, J. R. 2000. Thoughts on RFM scoring. *Journal of Database Marketing & Customer Strategy Management*, 8, 67–72.
- MORISADA, M., MIWA, Y., & DAHANA, W. D. 2019. Identifying valuable customer segments in online fashion markets: An implication for customer tier programs. *Electronic Commerce Research and Applications*, 33, 100822. <https://doi.org/https://doi.org/10.1016/j.elerap.2018.100822>
- MUTANDWA, E., KANUMA, N. T., RUSATIRA, E., KWIRINGIRIMANA, T., MUGENZI, P., GOVERE, I., & FOTI, R. 2009. Analysis of coffee export marketing in Rwanda: Application of the Boston consulting group matrix. *African Journal of Business Management*, 3(5), 210.
- NIDHRA, S., & DONDETI, J. 2012. Black box and white box testing techniques-a literature review. *International Journal of Embedded Systems and Applications (IJESA)*, 2(2), 29–50.
- PARASURAMAN, A. 1997. Reflections on gaining competitive advantage through customer value. *Journal of the Academy of Marketing Science*, 25(2), 154–161. <https://doi.org/10.1007/BF02894351>
- PUSPITASARI, N., HAVILUDDIN, H., & PUADI, F. U. J. H. 2022. Klasterisasi Wilayah Penghasil Tanaman Lada Menggunakan Algoritma K-Means. *Indonesian Journal of Computer Science*, 11(3).
- SAADEH, H., SAADEH, M., ALMOBAIDEEN, W., & AL-TAWIL, M. 2022. Evaluating the Optimal Number of Clusters to Identify Similar Gene Expression Patterns During Erythropoiesis. 2022 International Conference on Computer, Information and Telecommunication Systems (CITS), 1–5. <https://doi.org/10.1109/CITS55221.2022.9832988>
- SADIKU, M., SHADARE, A. E., MUSA, S. M., AKUJUOBI, C. M., & PERRY, R. 2016. Data visualization. *International Journal of Engineering Research And Advanced Technology (IJERAT)*, 2(12), 11–16.
- SARMAH, R. DAS, KALITA, J. K., & BHATTACHARYYA, D. K. 2011. A pattern matching approach for clustering gene expression data. *Int. J. Data Min. Model. Manag.*, 3, 130–149. <https://api.semanticscholar.org/CorpusID:10380267>
- SEVERIN, V., LOUVIERE, J. J., & FINN, A. 2001. The stability of retail shopping choices over time and across countries. *Journal of Retailing*, 77(2), 185–202.
- SHARYANTO, S., & LESTARI, D. 2022. Penerapan Data Mining Untuk Menentukan Segmentasi Pelanggan Dengan Menggunakan Algoritma K-Means dan Model RFM Pada E-Commerce. *JURIKOM (Jurnal Riset Komputer)*. <https://api.semanticscholar.org/CorpusID:255665256>
- SYAKUR, M. A., KHOTIMAH, B. K., ROCHMAN, E. M. S., & SATOTO, B. D. 2018. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conference Series: Materials Science and Engineering*, 336, 012017.
- WANG, Y., TANG, S., LIANG, F., ZHANG, Y., & LI, J. 2013. Beyond Kmedoids: Sparse Model Based Medoids Algorithm for Representative Selection. In S. Li, A. El Saddik, M. Wang, T. Mei, N. Sebe, S. Yan, R. Hong, & C. Gurrin (Eds.), *Advances in Multimedia Modeling* (pp. 239–250). Springer Berlin Heidelberg.
- ZAHARIA, S., & SCHMITZ, M. 2020. Customer Experience in Online-Retailing – An Analysis of the Main Segments in German Online-Retailing. In J. I. Kantola & S. Nazir (Eds.), *Advances in Human Factors, Business Management and Leadership* (pp. 177–188). Springer International Publishing.

Halaman ini sengaja dikosongkan