

METODE CLUSTERING DENGAN MODEL RFM DAN ALGORITMA K-MEANS UNTUK SEGMENTASI PELANGGAN TRANSJAKARTA

Trie Widiarti Ningsih, Sari Ningsih*

Sistem Informasi, Fakultas Teknologi Komunikasi Dan Informatika Universitas Nasional
Jalan Sawo Manila Pejaten Jakarta Selatan, Indonesia
sari.ningsih@civitas.unas.ac.id

ABSTRAK

Jakarta menghadapi tantangan besar dalam menyediakan transportasi umum yang efektif. Transjakarta sebagai moda angkutan massal berbasis sistem transit menjadi solusi utama dalam memenuhi kebutuhan mobilitas masyarakat. Namun, peningkatan kualitas layanan masih terkendala oleh kurangnya pemahaman terhadap perilaku pelanggan. Penelitian ini bertujuan untuk melakukan segmentasi pelanggan Transjakarta guna mendukung strategi pelayanan yang lebih tepat sasaran. Metode yang digunakan adalah model RFM (*Recency, Frequency, Monetary*) untuk mengukur perilaku pengguna, serta algoritma K-Means untuk mengelompokkan pelanggan berdasarkan kemiripan karakteristik. Jumlah cluster optimal ditentukan menggunakan metode *Elbow, Silhouette Score, dan Davies-Bouldin Index*. Hasil menunjukkan model tujuh cluster memberikan nilai *Silhouette Score* sebesar 0.789 dan *Davies-Bouldin Index* sebesar 0.303, yang menunjukkan kualitas pemisahan dan kepadatan cluster yang baik. Segmentasi menghasilkan tujuh kelompok pelanggan: Aktif, Rutin, Jarang, Sering Terjadi, Terbatas, Hemat, dan Minim. Dengan hasil ini, Transjakarta dapat menyusun strategi layanan yang lebih adaptif dan berkelanjutan, sesuai karakteristik dan kebutuhan tiap segmen pelanggan.

Kata kunci : Algoritma K-Means, Clustering, Model RFM, Segmentasi Pelanggan, Transjakarta.

1. PENDAHULUAN

Dalam era mobilitas perkotaan yang terus berkembang, transportasi umum memegang peran krusial dalam mendukung kehidupan masyarakat. Ditengah permasalahan yang semakin kompleks, Jakarta menghadapi tantangan besar dalam menyediakan layanan transportasi umum yang efektif dan efisien. Transportasi adalah bagian penting dari kehidupan masyarakat, tetapi sering menjadi sumber permasalahan seperti kemacetan lalu lintas, pencemaran udara, dan tingginya jumlah penduduk.

Sistem transportasi umum menjadi penting untuk mendukung mobilitas penduduk perkotaan. Transjakarta hadir sebagai solusi untuk memenuhi kebutuhan mobilitas penduduk Jakarta. Pada Januari 2004, Pemerintah Jakarta meluncurkan moda transportasi umum yang bernama Transjakarta. Transjakarta merupakan sistem transportasi umum yang menyediakan layanan berbasis transit dengan cepat, nyaman, dan terjangkau untuk memenuhi kebutuhan mobilitas penduduk di perkotaan. Transjakarta menawarkan jalur dan halte yang disediakan secara khusus, bersama dengan fasilitas bus yang nyaman dengan harga terjangkau dan kapasitas angkut yang luas. Di samping itu, Transjakarta juga mengoperasikan bus berwarna merah muda yang ditujukan khusus untuk penumpang wanita[1].

Namun, meskipun upaya Transjakarta untuk meningkatkan mobilitas perkotaan, masih terdapat sejumlah masalah yang perlu diatasi. Salah satu permasalahan utamanya adalah kesulitan dalam memahami perilaku pengguna Transjakarta secara menyeluruh. Dengan tingginya jumlah pengguna Transjakarta, diperlukan pemantauan terhadap

perilaku pengguna pada setiap periode guna meningkatkan kualitas pelayanan. Pemahaman yang mendalam terhadap perilaku pelanggan Transjakarta menjadi kunci strategis dalam meningkatkan kualitas layanan dan efisiensi operasional. Dalam upaya untuk mengoptimalkan pelayanan tersebut, metode segmentasi pelanggan menjadi semakin penting. Segmentasi pelanggan merupakan salah satu bagian yang krusial dalam mengidentifikasi dan mengelompokkan pelanggan berdasarkan perilaku pembelian, informasi demografis dan geografis, serta karakteristik psikografis[2]. Segmentasi pelanggan memungkinkan pihak manajemen untuk mengidentifikasi kelompok-kelompok pelanggan dengan karakteristik serupa, memahami kebutuhan mereka, dan merancang strategi pelayanan yang lebih terfokus.

Data mining adalah proses mencari pola dan informasi yang signifikan dalam kumpulan data besar. Saat mengimplementasikan data mining, berbagai teknik dapat diterapkan, diantaranya adalah unsupervised learning. Teknik dalam unsupervised learning yang relevan untuk segmentasi adalah *clustering*[3]. *Clustering* digunakan untuk mengidentifikasi dan mengelompokkan objek data yang memiliki kesamaan atau kearakteristik serupa[4]. *K-Means* adalah teknik *clustering* non-hierarchical yang membagi data menjadi beberapa *cluster* atau kelompok berdasarkan kemiripan di antara data. Data-data yang memiliki karakteristik serupa dikelompokkan ke dalam satu *cluster*, dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam *cluster* berdasarkan kemiripannya[5]. Maka salah satu pendekatan yang dapat diterapkan untuk penelitian ini

adalah memanfaatkan kombinasi model RFM (*Recency*, *Frequency*, *Monetary*) dan algoritma *K-Means*. Model RFM mengacu pada tiga aspek utama: *Recency*, *Frequency*, dan *Monetary*. *Recency* menunjukkan kapan pelanggan terakhir kali bertransaksi, *Frequency* menunjukkan seberapa sering pelanggan bertransaksi, *Monetary* mengacu pada total uang yang dihabiskan oleh pelanggan. Model RFM membantu dalam menemukan pelanggan yang paling mungkin merespons penawaran baru di antara pelanggan yang sudah ada [6]. Sementara algoritma *K-Means* untuk mengelompokkan nilai pelanggan ke dalam beberapa kelompok berdasarkan kemiripan. Dalam menentukan jumlah *cluster* yang paling optimal, metode elbow, *silhouette*, dan *davies bouldin index*, digunakan sebagai kunci dalam memastikan validitas hasil segmentasi.

2. TINJAUAN PUSTAKA

Beberapa penelitian sebelumnya telah mengeksplorasi penerapan teknik *clustering* untuk segmentasi pelanggan di berbagai sektor. Menurut [3] melakukan segmentasi pelanggan Telkomsel menggunakan metode *clustering* dengan model RFM dan algoritma *K-Means*. Penelitian ini berhasil mengidentifikasi empat segmen pelanggan dengan karakteristik berbeda, yaitu segmen pelanggan baru, potensial, loyal, dan yang memerlukan perhatian. Menurut [4] menerapkan Business Intelligence dengan teknik *K-Means Clustering* dan model RFM untuk segmentasi pelanggan di PT XYZ. Penelitian ini menunjukkan efektivitas sistem informasi Business Intelligence dalam memahami kontribusi pelanggan. Berdasarkan [5] melakukan segmentasi pelanggan PT Coversuper Indonesia Global dengan kombinasi model RFM dan *K-Means*, mengidentifikasi empat segmen pelanggan dan menunjukkan tingkat akurasi yang baik dalam karakteristik pelanggan. Menurut [15] menerapkan metode *K-Means* dan model RFM di PT Hasjrat Abadi Ambon, menunjukkan pembagian pelanggan ke dalam lima kelompok berdasarkan kebiasaan pembeli mereka. Dari [11] melakukan analisis RFM (*Recency*, *Frequency*, and *Monetary*) produk menggunakan metode *K-Means* pada Sintetic Store. Penelitian ini mengatasi penumpukan barang di gudang dengan mengelompokkan produk ke dalam tiga *cluster* utama: tertinggi, sedang, dan rendah. Hasil *clustering* memiliki kesesuaian 99% dengan perhitungan manual, dan sistem berfungsi baik di berbagai browser.

2.1. Segmentasi Pelanggan

Segmentasi pelanggan merupakan cara untuk membagi pasar menjadi kelompok-kelompok yang dapat dikelola berdasarkan karakteristik serupa untuk merancang strategi bisnis yang efektif. [7].

2.2. Data Mining dan Clustering

Data mining merupakan proses ekstraksi informasi penting dari data didalam database untuk menghasilkan informasi yang sangat berharga [8]. Data mining merupakan metode mengekstraksi informasi berguna dari data mentah berskala besar dan kemudian mengubahnya menjadi informasi yang membantu dalam mengambil keputusan yang lebih baik [9]. *Clustering* adalah salah satu teknik dari *unsupervised learning* yang dimanfaatkan untuk menganalisis data. [10]. *Clustering* adalah metode data mining tanpa arahan, artinya karakteristik setiap *cluster* didasarkan pada kesamaan atribut dari suatu *cluster*, bukan pada yang telah ditentukan sebelumnya [11]

2.3. RFM Model dan Algoritma K-Means

Teknik RFM sangat bergantung pada informasi transaksi yang tercatat oleh konsumen [12].

- Recency* menunjukkan bahwa pelanggan yang baru bertransaksi cenderung lebih responsif terhadap penawaran baru dibandingkan pelanggan yang sudah lama melakukan transaksi.
- Frequency* atau frekuensi menunjukkan bahwa pembelian mencerminkan seberapa sering konsumen melakukan pembelian. Jika pelanggan melakukan pembelian secara teratur, kemungkinan mereka akan merespons dengan lebih baik dari pada pelanggan yang jarang melakukan pembelian.
- Monetary* mengacu pada total uang yang dihabiskan pelanggan dalam pembelian mereka. Pelanggan yang mengeluarkan lebih banyak uang cenderung lebih responsif terhadap penawaran dari pada mereka yang menghabiskan jumlah yang lebih kecil [12].

K-Means adalah algoritma yang memerlukan parameter dan jumlah *cluster* sebagai input, dan membagi data dalam jumlah *cluster* yang telah ditentukan sebagai kesamaan di antara data dalam *cluster* tinggi [13][14].

2.4. Metode Elbow dan Silhouette

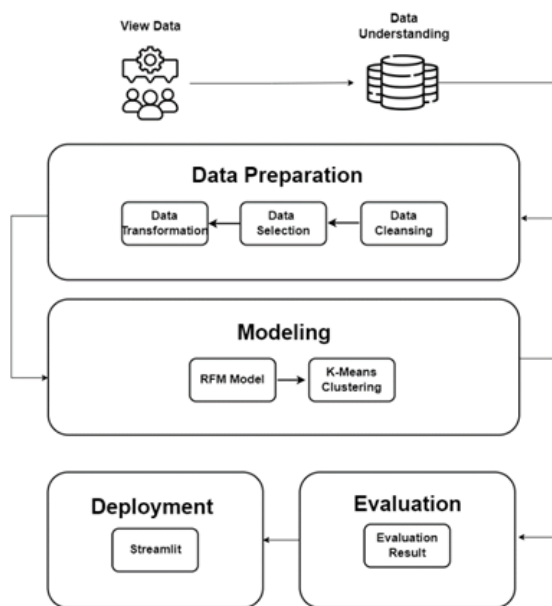
Tujuan utama dari metode elbow adalah memilih jumlah *cluster* yang optimal dengan mempertimbangkan tingkat kesalahan yang rendah [15] [16]. Penerapan metode elbow bertujuan untuk menentukan jumlah *cluster* optimal yang akan digunakan dalam proses *clustering* menggunakan algoritma *K-Means* [17] Metode *silhouette coefficient* merupakan kombinasi dari dua konsep, yaitu kohesion untuk mengukur kedekatan antara objek dalam suatu *cluster*, dan pemisahan untuk mengevaluasi perbedaan antara satu *cluster* dengan *cluster* lainnya [18]. Davies-Bouldin Index digunakan untuk menilai hasil pengelompokan dari suatu model *clustering*. DBI mengukur seberapa baik pengelompokan berdasarkan kedekatan data dengan titik pusat *cluster* dan jarak antara *centroid* dari masing masing *cluster*. Semakin kecil nilai DBI, semakin baik kualitas pengelompokan [18].

2.5. Google Colab, Rapid Miner dan Streamlit

Google Colab adalah platform yang digunakan untuk membuat, menjalankan, dan berbagi kode Python, serta melakukan pemrosesan data dan *machine learning*. Melalui Google Colab menyediakan lingkungan pengembangan yang nyaman dan terintegrasi dengan alat-alat populer seperti Jupyter Notebook, serta akses ke sumber daya komputasi yang kuat [19]. RapidMiner adalah sebuah perangkat lunak analisis data dan pengembangan data, yang dapat diintegrasikan dengan berbagai bahasa pemrograman dengan mudah. Ditulis menggunakan bahasa pemrograman java, RapidMiner dapat dijalankan pada berbagai sistem operasi [20]. Streamlit adalah sebuah *framework open source* berbasis python. Diciptakan dengan tujuan untuk membantu *developer* dalam membangun web interaktif dalam bidang ilmu data dan pembelajaran mesin atau *machine learning* [21].

3. METODE PENELITIAN

Penelitian ini menerapkan metode kuantitatif dengan menggunakan data transaksi Transjakarta dari bulan April 2023 yang diunduh dari Kaggle[22]. Tahapan penelitian berikut ini:



Gambar 1. Tahapan penelitian

Berdasarkan Gambar 1. Tahapan penelitiannya sebelum melakukan analisis data, dilakukan tahap data *preparation* untuk mempersiapkan data model, yaitu RFM. Proses segmentasi pelanggan menjadi kelompok dilakukan menggunakan metode *K-Means*, dengan menentukan jumlah *cluster* yang optimal berdasarkan metode Elbow, *Silhouette*, dan Davies-Bouldin Index. Metode ini memastikan bahwa hasil segmentasi dapat dipercaya dan akurat. Setelah proses *cluster* selesai, sebagaimana yang ditampilkan dalam

4. HASIL DAN PEMBAHASAN

4.1. Implementasi Tahapan Penelitian

Implementasi penelitian ini, meliputi view data, data understanding, data preparation, modeling, evaluation, dan deployment.

4.2. View Data

Tahapan memeriksa struktur dan jenis data dari dataset Transjakarta untuk memastikan bahwa data tersebut relevan digunakan dalam analisis segmentasi pelanggan.

4.3. Data Understanding

Dataset yang diunduh diimpor ke dalam Google Colab. Dataset berisi data transaksi pelanggan untuk bulan April 2023 dengan total 181.111 baris dan 22 kolom. Tipe data meliputi float 7 kolom, int 3 kolom, dan object 12 kolom.

4.4. Data Preparation

Proses untuk menyiapkan data akhir sebelum melakukan segmentasi dengan cara pembersihan missing value, pemilihan atribut, dan transformasi data sesuai kebutuhan.

4.5. Data Cleaning

Data *cleaning* dari nilai-nilai yang tidak lengkap atau null. Berikut adalah jumlah missing value pada dataset terlihat pada tabel 1:

Tabel 1. Missing Value

Atribut	Percentase (null)
transID	0.00
payCardID	0.00
payCardBank	0.00
payCardName	0.00
payCardSex	0.00
payCardBirthDate	0.00
corridorID	3.68
corridorName	7.16
direction	0.00
tapInStops	3.83
tapInStopsName	0.00

Tabel 1. Menunjukkan beberapa atribut dengan missing value sehingga harus dilakukan data *cleaning*. *Data Selection*

Tabel 2. Tabel Transaksi

	payCardName	tapOutTime	payAmount
0	User_0001	4/3/2023 7:13	3500
1	User_0001	4/4/2023 6:55	3500
2	User_0001	4/4/2023 18:15	3500
3	User_0001	4/5/2023 6:57	3500
4	User_0001	4/5/2023 18:54	3500
...	...	...	...
147008	User_0002	4/29/2023 11:53	3500

Tabel 2. Menunjukkan data transaksi sebanyak 147.008 pelanggan transjakarta yang sudah siap untuk proses selanjutnya .

4.6. Data Transformation

Tabel 3. Data Transformation

	payCardName	tapOutTime	payAmount
0	User_0001	4-3-2023 7:13	3500
1	User_0001	4-4-2023 6:55	3500
2	User_0001	4-4-2023 18:15	3500
3	User_0001	4-5-2023 6:57	3500
4	User_0001	4-5-2023 18:54	3500
...	...	...	...
147008	User_0002	2023-04-29	3500

Tabel 3. Menunjukkan data transformasi sebanyak 147.008 pelanggan transjakarta yang sudah siap untuk proses selanjutnya .

4.7. Modeling

Pada tahap modeling, dilakukan proses segmentasi pelanggan dengan menggunakan metode RFM (*Recency, Frequency, Monetary*) dan algoritma *K-Means clustering*. RFM dilakukan untuk mengelompokkan pelanggan berdasarkan nilai *Recency, Frequency*, dan *Monetary* dari data transaksi terlihat pada tabel 4.

Tabel 4. Data RFM

	payCardName	Recency	Frequency	Monetary
0	User_0001	27	36	126000
1	User_0008	26	36	0
2	User_0003	26	36	0
3	User_0005	25	36	0
4	User_0004	25	36	0
...	...	...	...	...
9163	User_0002	1	1	3500

Tabel 4. menunjukkan hasil dari model RFM, dengan informasi mengenai kapan pelanggan terakhir kali bertransaksi (*Recency*), seberapa sering pelanggan melakukan transaksi (*Frequency*), dan total uang yang pelanggan keluarkan (*Monetary*).

4.8. K-Means Clustering

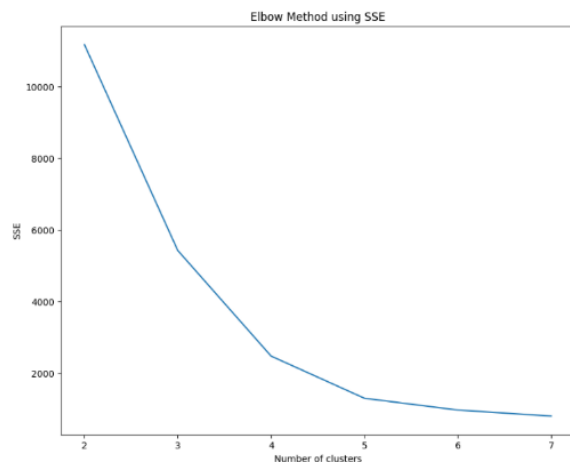
Pada tahap ini, atribut *Recency, Frequency*, dan *Monetary* distandarisasi menggunakan library Python scikit-learn dengan kelas *StandardScaler* untuk memastikan semua nilai memiliki skala yang seragam.

Tabel 5. Standarisasi

	Recency	Frequency	Monetary
0	0.871289	1.342088	2.054189
1	0.871289	1.409334	-0.635390
2	0.871289	1.005860	-0.635390
3	0.871289	1.207597	-0.635390
4	0.871289	0.938614	-0.635390
...	...	...	...
9163	-3.413830	-1.011510	-0.560679

Tabel 5. hasil standarisasi pada tahapan *K-Means Clustering* dengan RFM sebanyak 9.163 data.

Langkah selanjutnya adalah memilih nilai *k* yang optimal untuk jumlah *cluster* dengan memanfaatkan metode *elbow* dan *silhouette*. Berikut ini adalah hasil dari metode *elbow*:



Gambar 2. Metode Elbow

Berdasarkan Gambar 2 kurva Elbow, jumlah *cluster* optimal berada pada nilai 4, di mana grafik memiliki sudut yang paling tajam. Berikut adalah hasil metode *silhouette*:

Tabel 6. Silhouette Score

Jumlah Cluster	Silhouette Score
2	0.608014704705849
3	0.7154822674229963
4	0.7340031742934106
5	0.7415076576031576
6	0.7745118602138056
7	0.7890969140331999

Berdasarkan Tabel 6. *silhouette* menunjukkan jumlah *cluster* dari 2 sampai 8 yang tertinggi berada pada *cluster* 7 dan nilai 0.789 menunjukkan bahwa jumlah *cluster* yang paling optimal.

4.9. Evaluation

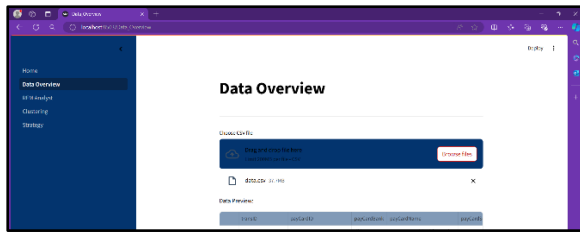
Tabel 7. Davies-Bouldin Index

Jumlah Cluster	davies-bouldin Score
2	0.7602915311885864
3	0.3594531663300993
4	0.3711130488662875
5	0.4815872815335568
6	0.40186400658727167
7	0.30261724306268073

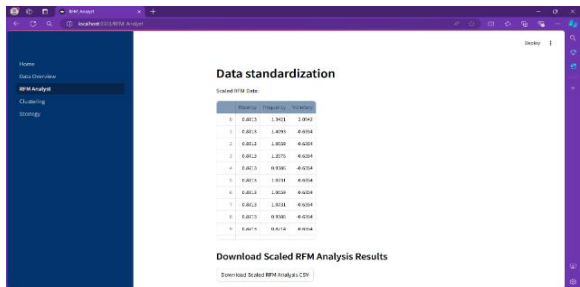
Berdasarkan Tabel 7. mengenai hasil Davies-Bouldin Index, jumlah *cluster* yang menunjukkan nilai terendah adalah 7 dengan nilai 0.303.

Deployment

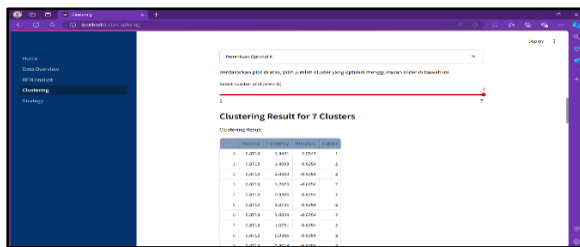
Pada tahap deployment, label *cluster* diintegrasikan dengan data asli untuk masing-masing entri, memungkinkan visualisasi data yang lebih informatif.



Gambar 3. Data overview



Gambar 4. RFM analyst



Gambar 5. Clustering

Berdasarkan gambar 3. data *overview*, dan gambar 4. RFM *analyst* menggambarkan segmentasi pelanggan Transjakarta. Visualisasi dilakukan menggunakan *framework* Streamlit, memanfaatkan model RFM dan teknik *clustering* K-Means untuk menghasilkan grafik dan tabel yang menggambarkan segmentasi pelanggan Transjakarta, terlihat pada gambar 5. merupakan hasil clustering secara visual.

4.10. Pengujian clustering

Metode *clustering* K-Means diuji menggunakan perangkat lunak RapidMiner untuk memvalidasi hasil *clustering* dan memastikan konsistensinya.

Row No.	label	Recency	Frequency	Monetary
1	cluster_1	0.871	1.342	2.054
2	cluster_2	0.871	1.409	-0.635
3	cluster_2	0.871	1.006	-0.635
4	cluster_2	0.871	1.208	-0.635
5	cluster_2	0.871	0.939	-0.635

Gambar 6. RapidMiner

Berdasarkan gambar 6. hasil *clustering* yang diperoleh dari pengujian menggunakan rapidminer, menggambarkan masing masing jenis clusternya.

Tabel 8. Jumlah Anggota Cluster

Cluster	Total
C0	421
C1	1983
C2	1741
C3	3076
C4	539
C5	921
C6	483

Berdasarkan Tabel 8. Mengenai jumlah anggota *cluster* yang menunjukkan jumlah cluster terbanyak di cluster C3 sebanyak 3076 pengguna trans.

4.11. Analisa Hasil

Tabel 9. Perbandingan Hasil Clustering

Cluster	Google Colab	Perhitungan Manual	RapidMiner
Cluster 0	421	421	421
Cluster 1	1983	1983	1983
Cluster 2	1741	1741	1741
Cluster 3	3076	3076	3076
Cluster 4	539	539	539
Cluster 5	921	921	921
Cluster 6	483	483	483

Berdasarkan Tabel 9. mengenai perbandingan hasil *cluster* yang menunjukkan hasil cluster dari ketiga metode menunjukkan hasil yang signifikan dari ke 7 cluster pengguna trans.

4.12. Visualisasi

Tabel 10. Rata-rata karakteristik

Cluster Label	Recency	Frequency	Monetary	Cluster
Jarang	17	11	41.000	Cluster 0
Aktif	27	33	116.716	Cluster 1
Rutin	27	33	62.320	Cluster 2
Sering Terjadi	22	3	5.486	Cluster 3
Terbatas	5	1	2.149	Cluster 4
Hemat	14	2	3.811	Cluster 5
Minim	17	12	40.702	Cluster 6

Berdasarkan Tabel 10. Visualisasi hasil *clustering* berdasarkan karakteristik RFM memberikan gambaran tentang distribusi data dan segmentasi pelanggan.

Tabel 11. Karakteristik

Cluster Label	Karakteristik
Jarang	Pelanggan dalam <i>cluster</i> ini menunjukkan nilai menengah untuk <i>Recency</i> dan <i>Monetary</i> , tetapi nilai rendah untuk <i>Frequency</i> .
Aktif	Pelanggan dalam <i>cluster</i> ini menunjukkan nilai tinggi untuk <i>Recency</i> , <i>Frequency</i> , dan <i>Monetary</i> .
Rutin	Pelanggan dalam <i>cluster</i> ini menunjukkan nilai tinggi untuk <i>Recency</i> dan <i>Frequency</i> , tetapi nilai menengah untuk <i>Monetary</i> .
Sering Terjadi	Pelanggan dalam <i>cluster</i> ini menunjukkan nilai menengah untuk <i>Recency</i> , dan nilai sangat rendah <i>Frequency</i> serta rendah untuk <i>Monetary</i> .
Terbatas	Pelanggan dalam <i>cluster</i> ini menunjukkan nilai sangat rendah untuk <i>Recency</i> , <i>Frequency</i> dan <i>Monetary</i> .
Hemat	Pelanggan dalam <i>cluster</i> ini menunjukkan nilai rendah untuk <i>Recency</i> dan sangat rendah untuk <i>Frequency</i> serta <i>Monetary</i> .
Minim	Pelanggan dalam <i>cluster</i> ini menunjukkan nilai menengah untuk <i>Recency</i> dan <i>Monetary</i> , dan nilai rendah untuk <i>Frequency</i> .

Berdasarkan Tabel 11. Tentang karakteristik meliputi kategori jarang, aktif, rutin, sering terjadi, terbatas, hemat dan minim.

Distribusi *Recency*: *Cluster* 1 dan 2 memiliki *Recency* tinggi (tidak melakukan transaksi baru-baru ini). *Cluster* 4 memiliki *Recency* rendah (baru melakukan transaksi).

Distribusi *Frequency*: *Cluster* 1 dan 2 memiliki *Frequency* tinggi (sering melakukan transaksi). *Cluster* 3, 4, dan 5 memiliki *Frequency* rendah (jarang melakukan transaksi).

Distribusi *Monetary*: *Cluster* 1 memiliki *Monetary* tinggi (menghabiskan banyak uang). *Cluster* 3, 4, dan 5 memiliki *Monetary* rendah (menghabiskan sedikit uang).

4.13. Rekomendasi Strategi

Berikut adalah rekomendasi strategi pemasaran berdasarkan karakteristik setiap *cluster*, diurutkan berdasarkan prioritas tinggi, menengah, dan rendah untuk memastikan strategi pemasaran yang lebih efisien dan terarah:

Tabel 12. Rekomendasi Strategi

Cluster Label	Jumlah Anggota	Rata-rata Umur	Jenis Kelamin	Rekomendasi Strategi
Aktif	1983	32.83	1059 F, 924 M	Promosi digital, program loyalitas, diskon bundling.
Rutin	1741	32.71	951 F, 790 M	Penawaran bundling, promosi spesial, referral program.
Hemat	921	32.51	495 F, 425 M	Diskon berkala, edukasi penghematan, promosi aplikasi, event khusus.
Sering Terjadi	3076	32.7	1639 F, 1437 M	Kampanye reaktivasi, survei kepuasan, penawaran khusus.
Terbatas	539	33.99	264 F, 275 M	Promosi musiman, edukasi rute dan layanan, insentif berbasis waktu.
Minim	438	33.57	252 F, 231 M 7.6	Promosi menarik, program pengenalan.
Jarang	421	33.49	220 F, 201 M	Kampanye kesadaran, fasilitas tambahan, penawaran pertama.

Berdasarkan Tabel 12. strategi pemasaran dibagi menjadi tiga prioritas utama berdasarkan karakteristik setiap *cluster* untuk memastikan pendekatan yang lebih efektif dan tepat sasaran. Prioritas Tinggi mencakup *Cluster* Aktif dan Rutin, dengan fokus utama pada peningkatan keterlibatan dan loyalitas pelanggan yang sudah sering menggunakan layanan. Prioritas Menengah meliputi *Cluster* Sering Terjadi dan Hemat, dengan tujuan utama meningkatkan frekuensi penggunaan dan menawarkan penawaran menarik bagi pelanggan. Prioritas Rendah mencakup *Cluster* Terbatas, Minim, dan Jarang, di mana strateginya bertujuan untuk menarik kembali pelanggan yang kurang aktif dan meningkatkan kesadaran tentang layanan. Dengan strategi yang diurutkan berdasarkan prioritas tinggi, menengah, dan rendah, diharapkan pendekatan pemasaran yang lebih

terarah dapat meningkatkan keterlibatan, loyalitas, dan kepuasan pelanggan Transjakarta.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengidentifikasi dan mengelompokkan pelanggan Transjakarta menjadi tujuh segmen menggunakan metode RFM dan algoritma *K-Means*. Setiap segmen memiliki karakteristik unik yang dapat digunakan untuk merumuskan strategi pemasaran yang lebih tepat sasaran. Tiga metode digunakan untuk evaluasi kualitas *cluster*, yaitu *Metode elbow*, *Silhouette Score*, dan *Davies-Bouldin Index*. Berdasarkan *Elbow Method*, jumlah *cluster* yang optimal adalah empat, karena pada titik ini penurunan Sum of Squared Error (SSE) mulai melandai. Skor *Silhouette* untuk model dengan tujuh *cluster* adalah 0.789, menunjukkan bahwa *cluster* tersebut memiliki pemisahan yang baik

dan kepadatan yang memadai. Nilai Davies-Bouldin Index untuk tujuh *cluster* adalah 0.303, menunjukkan bahwa model dengan jumlah *cluster* ini memiliki rasio antara jarak antar *cluster* dan jarak di dalam *cluster* yang optimal. Berdasarkan karakteristik ini, strategi pemasaran direkomendasikan dan dibagi menjadi tiga prioritas: tinggi, menengah, dan rendah. Prioritas tinggi ditujukan untuk segmen Aktif dan Rutin, dengan fokus pada promosi digital, program loyalitas, diskon bundling, penawaran bundling, promosi spesial, dan referral program. Prioritas menengah mencakup segmen Sering Terjadi dan Hemat, dengan strategi seperti kampanye reaktivasi, survei kepuasan, penawaran khusus, diskon berkala, edukasi penghematan, promosi aplikasi, dan event khusus. Prioritas rendah ditujukan untuk segmen Terbatas, Minim, dan Jarang, dengan strategi seperti promosi musiman, edukasi rute dan layanan, insentif berbasis waktu, promosi menarik, program pengenalan, kampanye kesadaran, fasilitas tambahan, dan penawaran pertama.

DAFTAR PUSTAKA

- [1] R. F. U. S. R. Sibarani and M. Megawati, "Pengaruh Kualitas Pelayanan dan Harga Terhadap Kepuasan Pelanggan Transjakarta," in *MDP Student Conference*, 2023, vol. 2, no. 2, pp. 567-573.
- [2] C. Carudin, "Utilization of Transaction Data for the Basis of Building A Strategy Based on Customer Characteristics with the *K-Means Clustering* Algorithm and RFM Model," *Jurnal Teknologi Terpadu*, vol. 7, no. 1, pp. 7-14, 2021.
- [3] M. A. Satriawan, R. Andreswari, and O. N. Pratiwi, "Segmentasi Pelanggan Telkomsel Menggunakan Metode *Clustering* Dengan RFM Model Dan Algoritma *K-Means*," *eProceedings of Engineering*, vol. 8, no. 2, 2021.
- [4] A. Karami and Y. Widharto, "Perancangan Business Intelligence Dan Segmentasi Pelanggan Menggunakan K Means *Clustering* Berdasarkan Rfm Model," *Industrial Engineering Online Journal*, vol. 12, no. 1, 2023.
- [5] A. T. Widiyanto and A. Witanti, "Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma *K-Means* Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global)," *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, vol. 1, no. 1, pp. 204-215, 2021.
- [6] İ. SABUNCU, E. TÜRKAN, and H. POLAT, "Customer segmentation and profiling with RFM analysis," *Turkish Journal of Marketing*, vol. 5, no. 1, pp. 22-36, 2020.
- [7] M. Harahap, Y. Lubis, and Z. Situmorang, "Analisis Pemasaran Bisnis dengan Data Science: Segmentasi Kepribadian Pelanggan berdasarkan Algoritma *K-Means Clustering*," *Data Sciences Indonesia (DSI)*, vol. 1, no. 2, pp. 76-88, 2021.
- [8] M. Fachry and F. R. Lubis, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA APIORI UNTUK REKOMENDASI JENIS ROTI BAGI PELANGGAN," *Djtechno: Jurnal Teknologi Informasi*, vol. 3, no. 1, pp. 102-107, 2022.
- [9] O. J. Chiedozie, M. I. Umar, and A. J. Chukwunwike, "Customer Segmentation using Data Mining Technique: A case study of Retail Industries in Nigeria," 2021.
- [10] P. Anitha and M. M. Patil, "RFM model for customer purchase behavior using *K-Means* algorithm," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 5, pp. 1785-1792, 2022.
- [11] R. Y. Firmansah, J. D. Irawan, and N. Vendyansyah, "Analisis Rfm (*Recency, Frequency and Monetary*) Produk Menggunakan Metode *K-Means*," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 5, no. 1, pp. 334-341, 2021.
- [12] B. C. Laksono and I. Y. Wulansari, "Pemodelan Dan Penerapan Metode Rfm Pada Estimasi Nilai Konsumen (Customer Lifetime Value) Menggunakan *K-Means Clustering* Machine Learning," *Semin. Nas. Off. Stat*, vol. 2020, no. 1, pp. 1277-1285, 2021.
- [13] M. Boentarman, S. Rostianingsih, and A. Setiawan, "Penerapan Segmentasi Pelanggan dengan Menggunakan Metode *K-Means Clustering* Pada Sistem Customer Relationship Management di PT. Titess," *Jurnal Infra*, vol. 9, no. 2, pp. 15-21, 2021.
- [14] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, "RFM ranking—An effective approach to customer segmentation," *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 10, pp. 1251-1257, 2021.
- [15] A. B. H. Kiat, Y. Azhar, and V. Rahmayanti, "Penerapan Metode *K-Means* Dengan Metode Elbow Untuk Segmentasi Pelanggan Menggunakan Model Rfm (*Recency, Frequency, & Monetary*)," *J. Repos*, vol. 2, no. 7, p. 945, 2020.
- [16] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah *Cluster* pada *K-Means Clustering*," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 14, no. 2, pp. 277-288, 2023.
- [17] N. T. Hartanti, "Metode Elbow dan *K-Means* Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional," *J. Nas. Teknol. dan Sist. Inf*, vol. 6, no. 2, pp. 82-89, 2020.
- [18] M. Sholeh and K. Aeni, "Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan *Silhouette* pada Model *Clustering* dengan Menggunakan Algoritma *K-Means*," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 8, no. 1, pp. 56-65, 2023.

- [19] R. G. Guntara, "Pelatihan Sains Data Bagi Pelaku UMKM di Kota Tasikmalaya Menggunakan Google Colab," *Joong-Ki: Jurnal Pengabdian Masyarakat*, vol. 2, no. 2, pp. 245-251, 2023.
- [20] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Regresi Linier," *Jurnal Nasional Teknologi dan Sistem Informasi (TEKNOSI)*, vol. 7, no. 1, pp. 8-17, 2021.
- [21] A. F. Hidayatullah, "Membuat Aplikasi Web Sains Data Dengan Mudah Menggunakan Streamlit," ed, 2021.
- [22] Dikisahkan, "Transjakarta - Public Transportation Transaction," Kaggle Dataset, April 2023. [Online]. Available: <https://www.kaggle.com/datasets/dikisahkan/transjakarta-transportation-transaction>.