



Analisis segmentasi pelanggan dengan model RFM (*Recency, Frequency, Monetary*) dan *K-Means Clustering* (Studi kasus: PT XYZ)

Ema Rosary Sitorus^{1✉}, Isna Nugraha¹

Program Studi S1 Teknik Industri, Universitas Pembangunan Nasional "Veteran" Jawa Timur Surabaya⁽¹⁾

DOI: [10.31004/jutin.v8i1.39447](https://doi.org/10.31004/jutin.v8i1.39447)

✉ Corresponding author:
[22032010177@student.upnjatim.ac.id]

Article Info	Abstrak
<p>Kata kunci: <i>K-Means Clustering;</i> <i>RFM;</i> <i>Segmentasi Pelanggan;</i> <i>Strategi Pemasaran;</i></p>	<p>Segmentasi pelanggan adalah proses krusial dalam memahami pola perilaku konsumen untuk mendukung pengambilan keputusan strategis dalam pemasaran. Tantangan utama yang dihadapi perusahaan adalah mengelompokkan pelanggan secara akurat berdasarkan data transaksi. Tujuan dilakukannya penelitian ini adalah untuk mengetahui dan melakukan segmentasi terhadap pelanggan menggunakan algoritma <i>K-Means clustering</i> berbasis model RFM (<i>Recency, Frequency, Monetary</i>) pada data transaksi penjualan AMDK di PT XYZ. Metode penelitian melibatkan analisis terhadap 111 data pelanggan yang diolah menggunakan <i>software</i> Orange Data Mining, dengan validasi hasil menggunakan <i>Silhouette Score</i> yang berguna dalam penentuan jumlah <i>cluster</i> ideal. Penelitian ini menghasilkan empat <i>cluster</i> pelanggan, dengan <i>Cluster 4</i> mencerminkan pelanggan dengan tingkat loyalitas tertinggi, ditandai oleh nilai <i>Frequency</i> dan <i>Monetary</i> yang dominan, sementara <i>Cluster 3</i> menggambarkan pelanggan dengan potensi loyalitas rendah. Hasil penelitian ini memberikan landasan ilmiah untuk pengembangan strategi pemasaran berbasis data yang lebih terfokus dan efisien.</p>
<p>Keywords: <i>Customer Segmentation;</i> <i>K-Means Clustering;</i> <i>Marketing Strategy;</i> <i>RFM;</i></p>	<p>Abstract</p> <p>Customer segmentation is a crucial process in understanding consumer behavior patterns to support strategic decision making in marketing. The main challenge companies face is accurately segmenting customers based on transaction data. The purpose of this research is to determine and segment customers using the <i>K-Means clustering</i> algorithm based on the RFM (<i>Recency, Frequency, Monetary</i>) model on AMDK sales transaction data at PT XYZ. The research method involves</p>

analysis of 111 customer data processed using Orange Data Mining software, with validation of the results using Silhouette Score which is useful in determining the ideal number of clusters. This research produces four customer clusters, with Cluster 4 reflecting customers with the highest level of loyalty, characterized by dominant Frequency and Monetary values, while Cluster 3 describes customers with low loyalty potential. The results of this research provide a scientific basis for the development of more focused and efficient data-based marketing strategies.

1. INTRODUCTION

Di tengah era digitalisasi dan persaingan yang semakin sengit, memahami perilaku konsumen secara mendalam menjadi suatu hal yang perlu diperhatikan bagi perusahaan agar tetap relevan, mempertahankan, serta memperluas pangsa pasar. Perilaku konsumen merujuk pada tindakan dan keputusan individu dalam proses pembelian maupun penggunaan produk atau layanan. Proses ini tidak hanya terbatas pada aktivitas pembelian, tetapi juga mencakup berbagai keputusan dan faktor yang memengaruhi keputusan tersebut (Setiawan et al., 2024). Segmentasi pasar merupakan proses pengelompokan pasar menjadi bagian-bagian yang lebih kecil, sementara segmentasi konsumen adalah upaya membagi target pasar berdasarkan karakteristik tertentu untuk merancang strategi bisnis yang lebih relevan (Awalina & Rahayu, 2023). Segmentasi ini bertujuan untuk mengelompokkan konsumen ke dalam subkategori yang unik, bermakna, dan homogen berdasarkan atribut serta karakteristik mereka. Dengan demikian, perusahaan dapat memahami pelanggan mereka secara lebih mendalam dan mengembangkan strategi yang efektif untuk menjalin hubungan sesuai dengan kebutuhan dan karakteristik setiap kelompok (Perdana, Satria et al., 2022).

Penelitian ini mengambil PT XYZ sebagai objek studi. PT XYZ memiliki berbagai jenis pelanggan, mulai dari individu rumah tangga hingga perusahaan dan institusi, dengan beragam latar belakang serta pola konsumsi. Meski memiliki basis pelanggan yang luas, perusahaan belum pernah menerapkan segmentasi pelanggan berbasis data secara formal. Strategi pemasaran yang digunakan masih bersifat umum dan tidak mempertimbangkan karakteristik atau preferensi pelanggan secara spesifik. Hal ini meningkatkan risiko kehilangan pelanggan, terutama jika kompetitor menawarkan sesuatu yang lebih menarik atau lebih sesuai dengan kebutuhan mereka (Fawzi et al., 2021). Dalam situasi persaingan yang semakin tajam, ketidakmampuan untuk mengelola segmen pelanggan dengan baik dapat menghambat perusahaan dalam mempertahankan pangsa pasar dan menghadapi tekanan dari pesaing. Maka dari itu, PT XYZ perlu menerapkan suatu pendekatan yang lebih terfokus untuk mengenali kebutuhan pelanggan serta merancang sebuah strategi pemasaran yang mampu menyesuaikan penawaran dan promosi dengan kebutuhan spesifik dari tiap-tiap segmen pelanggan (Yunus, 2016).

Untuk mengatasi masalah ini, PT XYZ perlu menerapkan segmentasi pelanggan menggunakan model RFM (*Recency, Frequency, Monetary*) berdasarkan tanggal terakhir pelanggan melakukan transaksi, frekuensi pelanggan melakukan transaksi, dan jumlah transaksi dalam periode tertentu (Taqwim et al., 2019). Metode *data mining* yang diterapkan untuk segmentasi pelanggan adalah *clustering*, dengan algoritma *K-Means* sebagai pendekatan utama. Data yang akan dianalisis menggunakan model RFM ialah data transaksi penjualan produk, setelah itu akan dilakukan proses segmentasi berdasarkan algoritma *K-Means* (Auliasari & Kertaningtyas, 2019). Jumlah *cluster* yang ideal ditentukan dengan indeks *Silhouette Score*. Gabungan model RFM dan metode *K-Means* mempermudah pengelompokan pelanggan ke dalam kategori yang tepat serta mengidentifikasi tingkat loyalitas mereka (Adiana et al., 2018). Melalui pendekatan ini, PT XYZ dapat mengidentifikasi pelanggan dengan risiko *churn* yang lebih tinggi, mengembangkan strategi retensi yang lebih efektif, dan menyusun strategi pemasaran yang lebih optimal dan menyesuaikan dengan setiap segmen pelanggan. Pada penelitian yang dilakukan oleh Wahyuni et al. (2023) berhasil menggabungkan model RFM dengan metode *K-Means* untuk segmentasi pelanggan di sektor percetakan. Hasilnya menunjukkan bahwa penggabungan kedua metode tersebut mampu menghasilkan segmentasi yang lebih akurat dan aplikatif. Namun, penelitian tersebut hanya menggunakan data transaksi pelanggan dalam periode tiga bulan, sehingga tidak mencakup pola fluktuasi musiman atau pengaruh promosi jangka panjang yang dapat memengaruhi perilaku pelanggan. Keterbatasan ini menjadi celah yang mendasari pentingnya penelitian lebih lanjut dengan cakupan data yang lebih panjang dan komprehensif.

Penelitian ini dilatarbelakangi oleh kebutuhan PT XYZ untuk memahami perilaku pelanggan secara lebih mendalam berdasarkan data penjualan selama 10 bulan terakhir, yakni dari Januari hingga Oktober 2024. Pemilihan periode waktu ini dapat mencakup pola pembelian yang stabil serta fluktuasi yang mungkin terjadi sepanjang tahun, termasuk periode promosi atau musim tertentu. Dengan data 10 bulan terakhir ini, PT XYZ dapat

memperoleh pemahaman yang lebih relevan dan terkini mengenai perilaku pelanggan, sehingga segmentasi yang dihasilkan dapat mencerminkan kondisi pasar yang aktual dan dinamis. Oleh karena itu, tujuan dilakukannya penelitian ini adalah untuk menganalisis segmentasi pelanggan dengan mengintegrasikan model RFM dan metode *K-Means* berdasarkan data penjualan terkini di PT XYZ. Segmentasi berbasis data yang dihasilkan akan memungkinkan PT XYZ mengidentifikasi segmen-segmen pelanggan dengan nilai jangka panjang yang tinggi, serta memahami kebutuhan unik dari masing-masing kelompok pelanggan. Dengan hasil analisis yang lebih spesifik ini, PT XYZ diharapkan mampu meningkatkan loyalitas pelanggan melalui strategi yang dipersonalisasi, sekaligus mencegah potensi kehilangan pelanggan bernilai tinggi. Lebih jauh, pemahaman segmentasi ini juga dapat memperkuat daya saing perusahaan dalam menghadapi kompetisi yang semakin ketat di industri bisnis, serta mendorong peningkatan profitabilitas dan kelancaran perusahaan di masa yang akan datang (Nursanti et al., 2024).

2. METHODS

2.1 Jenis dan Sumber Data

- a. Data kuantitatif merupakan data yang digunakan dalam penelitian ini, merupakan data yang dapat dihitung dan berbentuk nominal. Data transaksi penjualan AMDK termasuk dalam kategori data kuantitatif, karena berupa angka yang dapat diolah lebih lanjut untuk analisis.
- b. Sumber data yang digunakan dalam penelitian ini adalah data primer dan sekunder. Data primer dikumpulkan langsung dari sumber penelitian, yaitu transaksi pelanggan yang tercatat pada sistem penjualan (Bykova et al., 2024). Data tersebut dikumpulkan untuk dijadikan dataset yang akan dianalisis menggunakan metode data mining.

2.2 Metode Pengumpulan Data

- a. Wawancara
Dalam melaksanakan tahap ini dilakukan dengan mengajukan pertanyaan langsung kepada pihak yang relevan, dalam hal ini penulis melakukan wawancara kepada unit penjualan, untuk mendapatkan informasi terkait dengan proses transaksi pelanggan (Wildemuth, 2017).
- b. Observasi
Metode observasi digunakan untuk memperoleh data melalui pengamatan langsung pada lokasi objek penelitian, terutama pada proses bisnis yang terjadi di bagian penjualan.
- c. Dokumentasi
Dokumentasi digunakan untuk mengumpulkan catatan atau dokumen hasil wawancara dan observasi, yang kemudian digunakan dalam analisis penelitian.
- d. Studi Kepustakaan
Pada tahap ini peneliti menelaah referensi terkait, yaitu penelitian sebelumnya yang relevan dengan topik segmentasi pelanggan menggunakan *data mining* dan algoritma *K-Means clustering*.

2.3 Data Mining Life Cycle

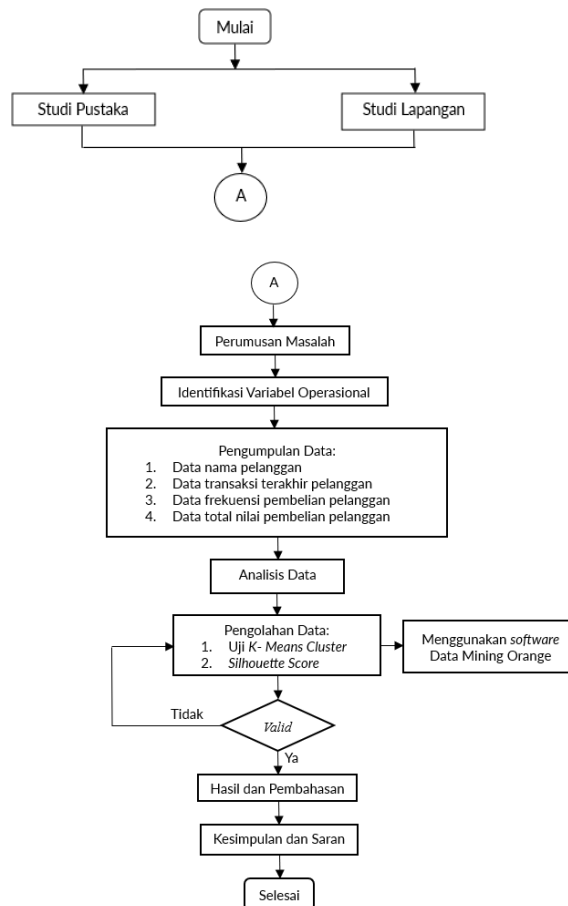
- a. *Data Cleaning* (Pembersihan Data)
Data yang terkumpul diproses untuk menghilangkan data yang tidak relevan atau data yang hilang agar analisis dapat dilakukan dengan data yang bersih dan akurat (Osborne, 2013).
- b. *Data Integration* (Integrasi Data)
Data yang telah terkumpul dari berbagai sumber dari berbagai sumber digabungkan menjadi *dataset* untuk memastikan bahwa semua informasi yang diperlukan terkumpul (Hermawan et al., 2024).
- c. *Data Selection* (Seleksi Data)
Proses memilih dan memilah data yang relevan untuk dianalisis, dalam penelitian ini yaitu data transaksi yang berhubungan dengan variabel *Recency*, *Frequency*, dan *Monetary* (Reyes, 2020).
- d. *Data Transformation* (Transformasi Data)
Data diubah menjadi format yang sesuai dengan prosedur analisis yang akan dilakukan menggunakan algoritma *K-Means clustering*.
- e. *Data Mining* (Penggalian Data)
Pada tahap ini, algoritma *K-Means clustering* diterapkan untuk dilakukan segmentasi pelanggan ke dalam beberapa kelompok berdasarkan tingkat kesamaan mereka. Dalam penelitian ini, *software* Orange Data Mining digunakan untuk melakukan analisis *clustering*.

f. *Patterns Evaluation* (Evaluasi Pola)

Hasil *clustering* dievaluasi dengan menggunakan metode *Silhouette Score* untuk menentukan kualitas pengelompokan yang dihasilkan dan memastikan bahwa jumlah *cluster* yang dibuat adalah yang optimal.

2.4 Proses Algoritma K-Means

- Penentuan jumlah kelompok *clustering* K yang diinginkan.
- Menginisiasi pusat *cluster* secara acak.
- Melakukan perhitungan jarak setiap data terhadap pusat *cluster* menggunakan rumus *Euclidean Distance*.
- Mengklasifikasikan data ke dalam *cluster* berdasarkan jarak terdekat ke pusat *cluster*.
- Memperbarui posisi pusat *cluster* dengan menghitung rata-rata data dalam *cluster* tersebut.
- Mengulangi langkah 3 hingga 5 sampai posisi pusat *cluster* tidak berubah lagi, yang menunjukkan bahwa proses *clustering* telah konvergen (Trianasari & Permadi, 2024).

2.5 Flowchart

Gambar 1. Flowchart

Gambar 1 *Flowchart* adalah representasi visual dari langkah-langkah atau proses dalam bentuk diagram yang terdiri dari simbol-simbol tertentu. Dalam konteks ini, *flowchart* bertujuan untuk mengilustrasikan tahapan penyelesaian penelitian dari awal sampai selesai. Penelitian diawali dengan kajian pustaka dan observasi lapangan, kemudian dilanjutkan dengan merumuskan masalah dan mengidentifikasi variabel operasional. Selanjutnya, dilakukan pengumpulan data pelanggan yang mencakup informasi nama, transaksi terakhir, frekuensi pembelian, serta total nilai pembelian. Data yang terkumpul dianalisis menggunakan perangkat lunak Data Mining Orange dengan pendekatan *K-Means Cluster* dan *Silhouette Score* untuk mengevaluasi validitas hasilnya. Apabila hasilnya valid, langkah berikutnya adalah pembahasan serta penyusunan kesimpulan dan saran. Proses penelitian berakhir setelah seluruh langkah diselesaikan.

3. RESULT AND DISCUSSION**3.1 Segmentasi Pelanggan**

Segmentasi merupakan kegiatan pembagian pelanggan ke dalam beberapa kelompok berdasarkan tingkat loyalitas pelanggan tersebut untuk merancang strategi pemasaran yang efektif. Ini merupakan langkah pertama

dalam mengembangkan model bisnis. Segmentasi dilakukan dengan mempertimbangkan 6 karakteristik yang diperoleh dari nilai RFM, seperti pada tabel di bawah ini (Adiana et al., 2018).

Tabel 1. Karakteristik Pelanggan

Kelas Pelanggan	Karakteristik
<i>Superstar</i>	a. Pelanggan dengan loyalti yang tinggi
	b. Mempunyai nilai <i>monetary</i> yang tinggi
	c. Mempunyai frekuensi yang tinggi
	d. Mempunyai transaksi paling tinggi.
<i>Golden Customer</i>	a. Mempunyai nilai <i>monetary</i> tertinggi yang ke dua
	b. <i>Frequency</i> yang tinggi
	c. Mempunyai rata-rata transaksi.
<i>Typical Customer</i>	Mempunyai rata-rata nilai <i>monetary</i> dan rata-rata transaksi
<i>Occational Customer</i>	a. Nilai <i>monetary</i> terendah kedua setelah <i>dormant customer</i>
	b. Nilai <i>recency</i> paling rendah
	c. Transaksi paling tinggi
Everyday Shopper	a. Memiliki peningkatan transaksi
	b. Transaksi yang rendah
	c. Mempunyai nilai <i>monetary</i> sedang sampai dengan rendah
Dormant Customer	a. Mempunyai <i>frequency</i> dan <i>monetary</i> yang paling rendah
	b. Nilai <i>recency</i> yang paling rendah.

Segmentasi pelanggan bertujuan untuk membagi pelanggan ke dalam kelompok dengan kebutuhan dan perilaku yang serupa. Setiap kelompok diharapkan memiliki kesamaan dalam satu aspek tertentu, seperti nilai mereka bagi organisasi atau pola interaksi mereka terhadap produk atau layanan. Dengan segmentasi ini, organisasi dapat menyusun strategi pemasaran yang lebih terarah untuk setiap kelompok. Aktivitas pemasaran, seperti komunikasi, penawaran produk atau layanan, serta program yang dirancang, menjadi lebih efektif karena disesuaikan dengan kesamaan dalam kebutuhan dan perilaku tiap segmen (Hananto et al., 2017).

3.2 Pengumpulan dan Pembersihan Data

Transaksi dalam periode tertentu dikumpulkan untuk dianalisis. Dengan memanfaatkan data rekap transaksi *online* dari *website* perusahaan dalam jangka waktu Januari hingga Oktober 2024 yang diperoleh dari divisi penjualan AMDK. Proses ini dilakukan agar data atau atribut yang tidak tersedia dapat dihilangkan (Shaliha et al., 2021). Dapat dilihat pada Gambar 2. Data Transaksi AMDK di PT XYZ Sebanyak 2.845 rekap transaksi berhasil dikumpulkan dalam periode tersebut.

	A	B	C
1	Nama Toko	Satuan	Harga
2	Pembelian_1	20	297.297,20
3	Pembelian_2	120	153.189,60
4	Pembelian_3	21	283.783,71
5	Pembelian_4	25	337.837,75
2842	Pembelian_2841	15	236.486,55
2843	Pembelian_2842	20	315.315,40
2844	Pembelian_2843	25	360.360,25
2845	Pembelian_2844	23	331.531,43

Gambar 2. Data Transaksi AMDK di PT XYZ

3.3 Integrasi Data

Pada tahap ini penggabungan data transaksi dari pelanggan yang sama serta pelanggan yang melakukan transaksi lebih dari satu kali, merupakan tahap dari proses integrasi data (Rizki et al., 2019). Sebelum penggabungan, penulis terlebih dahulu mencatat informasi penting, seperti tanggal transaksi terakhir yang dilakukan oleh setiap pelanggan, jumlah total transaksi yang dikeluarkan oleh setiap pelanggan, serta total nilai belanja yang terkumpul selama periode analisis, yaitu dari bulan Januari hingga Oktober 2024. Langkah ini bertujuan untuk memastikan data yang digabungkan mencerminkan aktivitas transaksi pelanggan secara akurat, yang dapat dilihat pada Gambar 3. Transaksi dari Pelanggan yang Sama.

	A	B	C
3	Nama Toko	Satuan	Harga
4	Customer_1	10	Rp 157.657,70
5	Customer_1	15	Rp 236.486,55
6	Customer_1	15	Rp 202.702,65
7	Customer_1	15	Rp 222.972,90
8	Customer_1	17	Rp 245.044,97
9	Customer_1	17	Rp 245.044,97

Gambar 3. Transaksi dari Pelanggan yang Sama

3.4 Seleksi Data

Peneliti memilih data yang dianggap relevan dan dapat dijadikan sebagai variabel untuk dianalisis (Kusumo et al., 2019). Pada Gambar 4. Seleksi Rekap Data Transaksi dapat dilihat atribut yang akan digunakan dalam penelitian ini diambil dari atribut nama pelanggan dan jumlah transaksi.

	A	B
3	Nama Toko	Harga
4	Customer_1	Rp 157.657,70
5	Customer_1	Rp 236.486,55
6	Customer_1	Rp 202.702,65
7	Customer_1	Rp 222.972,90
8	Customer_1	Rp 245.044,97
9	Customer_1	Rp 245.044,97

Gambar 4. Seleksi Rekap Data Transaksi

3.5 Transformasi Data

Pada tahap ini, data diubah menjadi format yang siap untuk diproses lebih lanjut dalam data mining. Penulis menambahkan tiga atribut baru, yaitu *Recency* (hari), *Frequency* (jumlah transaksi), dan *Monetary* (total belanja), yang digunakan untuk membangun model RFM (Gustriansyah et al., 2020).

- Recency* dihitung dengan mengukur selisih waktu antara tanggal terakhir pelanggan melakukan transaksi dan tanggal penelitian, yang menggambarkan waktu yang telah berlalu sejak transaksi terakhir tersebut.
- Frequency* menggambarkan berapa kali pelanggan melakukan transaksi selama periode analisis.
- Monetary* mengacu pada total pengeluaran pelanggan sepanjang periode yang dianalisis, dihitung dengan menjumlahkan seluruh pembayaran yang dilakukan pelanggan (Sabuncu et al., 2020).

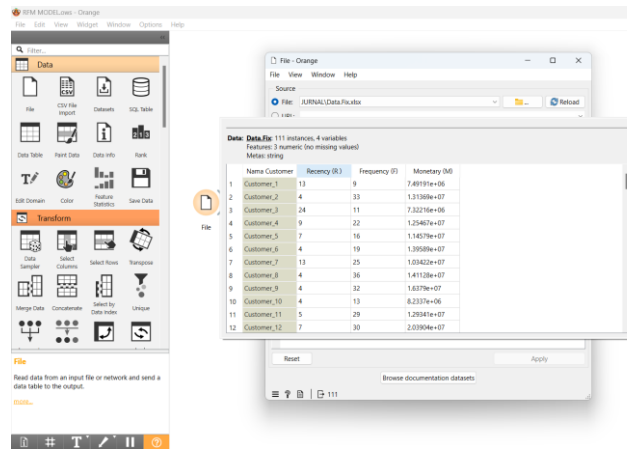
Setelah itu, dilakukan penggabungan data transaksi untuk mengurangi redundansi. Dapat dilihat pada Gambar 5. Pembuatan Atribut *Recency*, *Frequency*, dan *Monetary*, data yang semula berjumlah 2845 transaksi, di mana beberapa pelanggan tercatat lebih dari sekali, setelah digabungkan menjadi 111 transaksi yang mewakili 111 pelanggan unik. Proses penggabungan ini bertujuan untuk menghindari pencatatan ganda dan memastikan bahwa setiap data transaksi yang digunakan mewakili satu pelanggan yang berbeda, sehingga analisis *clustering* yang dilakukan dapat menghasilkan pemodelan yang lebih akurat dan optimal.

	A	B	C	D
1	Nama Customer	Recency (R)	Frequency (F)	Monetary (M)
2	Customer_1	13	9	7.491.908,16
3	Customer_2	4	33	13.136.937,42
4	Customer_3	24	11	7.322.159,48
5	Customer_4	9	22	12.546.666,66
109	Customer_108	4	46	13.844.197,47
110	Customer_109	13	20	11.587.450,42
111	Customer_110	7	30	9.334.381,46
112	Customer_111	13	19	8.722.036,46

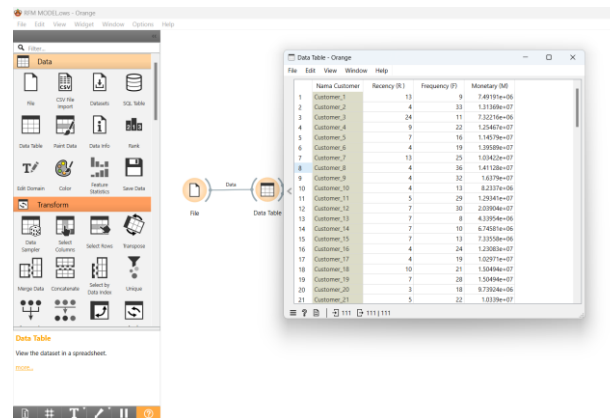
Gambar 5. Pembuatan Atribut *Recency*, *Frequency*, dan *Monetary*

3.6 Penggalan Data

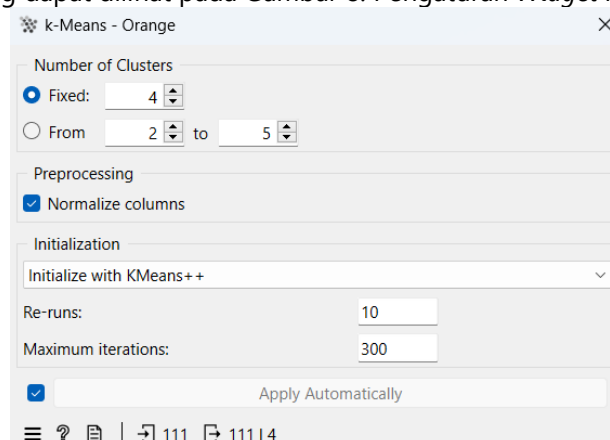
Dalam penelitian ini, teknik *data mining* diterapkan menggunakan algoritma *K-Means clustering*, dengan model yang diimplementasikan lewat perangkat lunak Orange Data Mining. Gambar 6. Proses *Import Data* merupakan tahap pertama yang harus dilakukan, yaitu peneliti menambahkan *Widget File* ke dalam lembar kerja dan mengimpor *file* data yang akan digunakan untuk penelitian ke dalam *software* Orange Data Mining.

Gambar 6. Proses *Import Data*

Langkah selanjutnya dapat dilihat pada Gambar 7. Proses *Priview Data*, pada tahapan ini peneliti menambahkan *Widget Data Table* guna memeriksa pratinjau data yang akan diproses dan memastikan bahwa tidak ada data yang hilang.

Gambar 7. Proses *Priview Data*

Setelah memverifikasi bahwa data lengkap, peneliti melanjutkan dengan menambahkan *widget K-Means* yang dihubungkan dengan *Widget Data Table* untuk melakukan proses klusterisasi menggunakan algoritma *K-Means*. Atribut yang digunakan dalam klusterisasi ini meliputi *Recency*, *Frequency*, dan *Monetary*. Pada tahap ini, peneliti memilih pengaturan *K-Means+* dan mengaktifkan opsi *maximum iterations* dengan batas iterasi maksimum sebanyak 300 yang dapat dilihat pada Gambar 8. Pengaturan *Widget K-Means*.

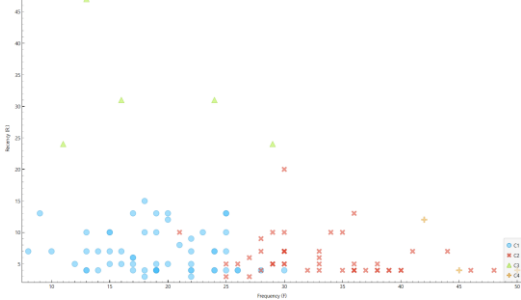
Gambar 8. Pengaturan *Widget K-Means*

Setelah proses klusterisasi pada Gambar 8. Pengaturan *Widget K-Means* diatas dengan pengelompokan data menjadi 4 kategori *cluster*, diperoleh hasil kelompok (*cluster*) 1 berisi 57 data, kelompok (*cluster*) 2 berisi 46 data, kelompok (*cluster*) 3 berisi 5 data, dan kelompok (*cluster*) 4 berisi 3 data. Informasi lengkap mengenai keanggotaan setiap *cluster* dapat dilihat pada Tabel 2. Keanggotaan Setiap *Cluster* yang menunjukkan keanggotaan masing-masing *cluster*.

Tabel 2. Tabel Keanggotaan Setiap *Cluster*

Cluster	Anggota Cluster
Cluster 1	1, 4, 5, 6, 7, 10, 13, 14, 15, 16, 17, 20, 21, 24, 28, 29, 30, 31, 32, 33, 35, 38, 39, 41, 42, 43, 44, 45, 48, 49, 52, 56, 57, 58, 60, 61, 62, 63, 67, 70, 73, 74, 75, 77, 82, 85, 86, 87, 92, 94, 97, 101, 103, 104, 106, 109, 111. (57)
Cluster 2	2, 8, 9, 11, 12, 18, 19, 22, 23, 25, 26, 27, 34, 36, 37, 40, 46, 47, 50, 51, 53, 54, 55, 64, 65, 66, 68, 68, 71, 76, 77, 79, 80, 81, 84, 88, 89, 90, 91, 95, 96, 99, 100, 102, 108, 110. (46)
Cluster 3	3, 83, 93, 105, 107. (5)
Cluster 4	59, 72, 98. (3)

Kemudian hasil klasterisasi disajikan menggunakan 3 metode yakni *Scatter Plot* pada gambar 9, *Silhouette Plot* pada gambar 10, dan *Box Plot* pada gambar 11, 12, dan 13.

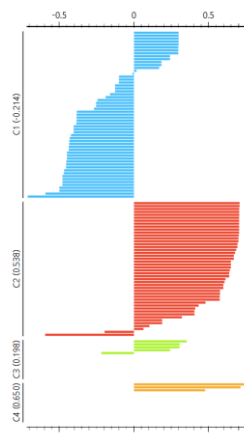


Gambar 9. Scatter Plot Hasil Cluster

Berdasarkan hasil visualisasi *scatter plot clustering* menggunakan model RFM di Gambar 9. *Scatter Plot* Hasil Cluster, data pelanggan dikelompokkan ke dalam beberapa cluster yang merepresentasikan karakteristik perilaku pelanggan. Pada sumbu X (*Frequency*), data mencerminkan seberapa sering pelanggan melakukan pembelian, sementara sumbu Y (*Recency*) menunjukkan waktu sejak transaksi terakhir. Hasil analisis menunjukkan bahwa setiap cluster memiliki pola yang unik.

Cluster pertama (C1) yang didominasi oleh lingkaran biru berada pada area dengan frekuensi rendah hingga menengah serta *recency* rendah. Hal ini menunjukkan bahwa pelanggan dalam cluster ini aktif bertransaksi baru-baru ini namun dalam intensitas pembelian yang tidak terlalu sering. Cluster kedua (C2) yang direpresentasikan oleh simbol tanda silang merah menunjukkan pelanggan dengan frekuensi rendah tetapi memiliki variasi *recency* yang lebih tinggi. Ini mengindikasikan bahwa pelanggan dalam cluster ini melakukan pembelian yang jarang dengan pola waktu transaksi yang tersebar lebih luas. Sementara itu, cluster ketiga (C3) yang diwakili oleh segitiga hijau terlihat berada pada area dengan frekuensi tinggi dan *recency* tinggi. Pelanggan dalam cluster ini memiliki pola pembelian yang signifikan namun sudah cukup lama tidak melakukan transaksi terbaru. Cluster ini menjadi perhatian khusus untuk strategi reaktivasi pelanggan. Selain itu, cluster lainnya seperti C4 (simbol "+") memperlihatkan distribusi yang unik dan dapat mencerminkan segmen dengan karakteristik tertentu, seperti loyalitas tinggi atau kontribusi signifikan terhadap pendapatan.

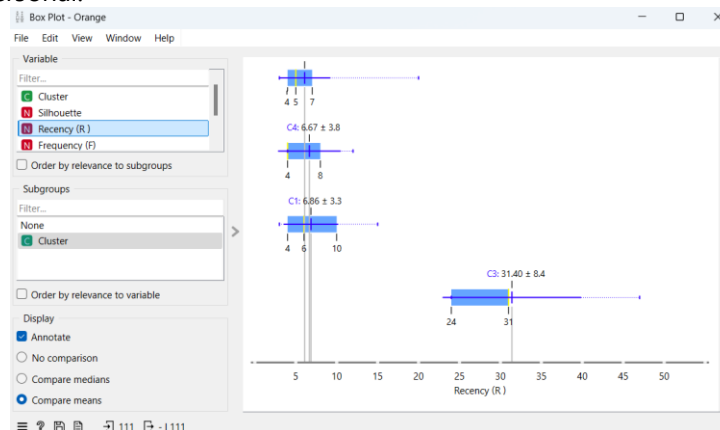
Adanya *outlier* di beberapa cluster, seperti titik-titik yang berada jauh dari pusat cluster, juga mengindikasikan bahwa terdapat pelanggan dengan perilaku pembelian yang sangat berbeda dibandingkan pelanggan lainnya. Hal ini bisa menjadi peluang untuk menyusun strategi personalisasi khusus. Secara keseluruhan, hasil *clustering* ini membantu PT XYZ memahami perilaku pelanggan berdasarkan model RFM dan mengidentifikasi segmen-segmen utama. Strategi pemasaran yang lebih terarah, seperti program loyalitas untuk cluster dengan frekuensi tinggi atau kampanye reaktivasi untuk cluster dengan *recency* tinggi, dapat diterapkan untuk meningkatkan kinerja penjualan.



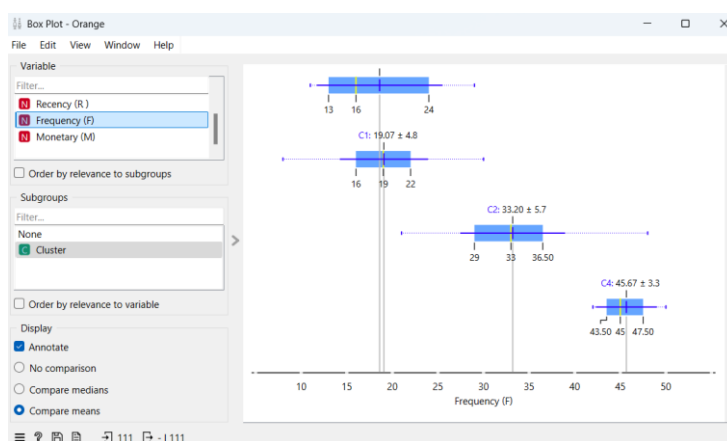
Gambar 10. Silhouette Plot Hasil Clustering

Hasil analisis menggunakan *silhouette plot* pada Gambar 10. *Silhouette Plot Hasil Clustering* diatas menunjukkan kualitas pemisahan antar *cluster* dalam segmentasi pelanggan berdasarkan model RFM dan metode *K-Means*. *Plot* ini mengilustrasikan nilai *silhouette* untuk setiap data pada masing-masing *cluster*, yang mencerminkan tingkat kohesi internal *cluster* dan separasi antar *cluster*. *Cluster* 1 (C1), yang ditandai dengan batang berwarna biru, memiliki nilai *silhouette* positif yang cukup tinggi dan konsisten. Hal ini menunjukkan bahwa anggota dalam *cluster* ini terkelompok dengan baik, dengan jarak *intra-cluster* yang kecil dan jarak *inter-cluster* yang signifikan terhadap *cluster* lainnya. *Cluster* ini dapat diinterpretasikan sebagai kelompok pelanggan dengan karakteristik yang homogen. *Cluster* 2 (C2), yang direpresentasikan dengan batang merah, juga memiliki nilai *silhouette* positif yang cukup baik. Namun, terdapat variasi nilai yang lebih besar dibandingkan C1, menunjukkan bahwa beberapa anggota *cluster* ini memiliki jarak yang lebih dekat ke *cluster* lain. Hal ini mengindikasikan kemungkinan adanya pelanggan di C2 yang memiliki karakteristik mirip dengan pelanggan dari *cluster* lain. Sementara itu, *cluster* kecil seperti C3 (kuning) dan C4 (hijau) memiliki batang dengan panjang yang lebih pendek dan nilai *silhouette* yang bervariasi. *Cluster* ini tampaknya berisi kelompok pelanggan dengan jumlah data yang lebih sedikit dan karakteristik yang lebih unik. Meskipun mereka memiliki nilai *silhouette* positif, perlu diperhatikan bahwa kohesi *internal cluster* ini mungkin tidak sekuat *cluster* lainnya.

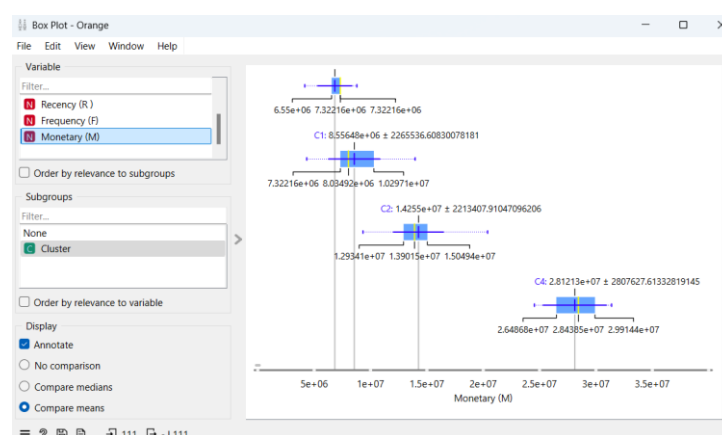
Secara keseluruhan, sebagian besar nilai *silhouette* positif menunjukkan bahwa metode *clustering* telah menghasilkan pemisahan yang cukup baik di antara segmen pelanggan. Namun, beberapa anggota *cluster* dengan nilai *silhouette* mendekati nol atau negatif perlu dianalisis lebih lanjut karena mereka menunjukkan kemungkinan tumpang tindih antar *cluster*. Hal ini memberikan peluang untuk melakukan evaluasi ulang terhadap jumlah optimal *cluster* atau parameter *clustering* guna meningkatkan akurasi segmentasi pelanggan. Hasil ini memberikan wawasan strategis bagi PT XYZ untuk mengembangkan pendekatan yang lebih terarah terhadap pelanggan berdasarkan segmen mereka, dengan fokus pada peningkatan pengalaman pelanggan dan strategi pemasaran yang lebih personal.



Gambar 11. Box Plot Hasil Atribut Recency



Gambar 12. Box Plot Hasil Atribut Frequency



Gambar 13. Box Plot Hasil Atribut Monetary

Tabel 3. Nilai Centroid Setiap Cluster

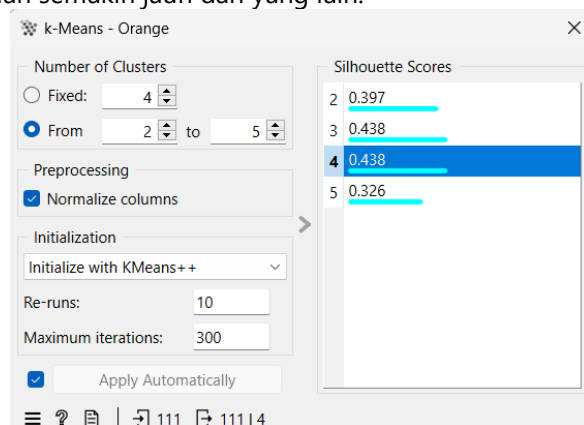
Cluster	Recency	Frequency	Monetary	Rank
C1	6,86	19,07	Rp. 8.556.480	3
C2	6,00	33,20	Rp. 14.255.000	2
C3	31,40	18,60	Rp. 6.829.230	4
C4	6,67	45,67	Rp. 28.121.300	1

Berdasarkan hasil yang tercantum dalam Tabel 3, *Cluster 4* adalah kelompok pelanggan dengan peringkat tertinggi, dengan nilai rata-rata recency sebesar 6,67; frekuensi sebesar 45,67 dan nilai moneter sebesar Rp. 28.121.300. *Cluster 2*, yang berada di peringkat kedua, memiliki nilai rata-rata variabel kecenderungan sebesar 6,00; frekuensi sebesar 33,20 dan nilai moneter sebesar Rp. 14.255.000. Sementara itu, *Cluster 1* yang berada di peringkat ketiga, memiliki nilai frekuensi rata-rata sebesar 6,86; frekuensi rata-rata sebesar 19,07 dan nilai moneter sebesar Rp. 8.556.480. Terakhir, *Cluster 3*, yang berada di peringkat keempat, memiliki nilai frekuensi rata-rata sebesar 31,40; frekuensi rata-rata sebesar 18,60 dan nilai moneter sebesar Rp. 6.829.230. *Clustering* dengan model RFM menghasilkan segmentasi pelanggan pada setiap *cluster* dari 111 data transaksi penjualan AMDK yang dianalisis. Dari model RFM tersebut, pelanggan yang termasuk dalam kategori sangat loyal terdapat di *Cluster 4*, dengan 3 pelanggan. Ini disebabkan oleh fakta bahwa mereka memiliki nilai RFM lebih rendah daripada pelanggan di *Cluster 4*, dan *Cluster 2*, yang terdiri dari 46 pelanggan, dianggap sebagai pelanggan loyal karena memiliki nilai RFM yang sebanding dengan pelanggan di *Cluster 4*. *Cluster 3*, yang terdiri dari 5 pelanggan, dianggap sebagai pelanggan tidak potensial karena memiliki nilai RFM yang sangat jauh dari nilai pelanggan *Cluster 4*.

3.7 Evaluasi Pola

Tahap terakhir adalah *evaluation* (evaluasi pola), yaitu proses evaluasi terhadap data yang telah divisualisasikan atau dimodelkan untuk menilai pola-pola yang dihasilkan, memastikan bahwa pola tersebut sesuai dengan tujuan analisis, serta mengidentifikasi sejauh mana pola-pola tersebut memberikan *insight* yang relevan, akurat, dan aplikatif dalam mendukung pengambilan keputusan atau penyusunan strategi yang efektif (Sastya & Nugraha, 2023). Dalam penelitian ini, evaluasi terhadap hasil klasterisasi dilakukan menggunakan metode

Silhouette Score. Tujuan dari pemeriksaan ini adalah untuk mengetahui jumlah kluster yang dipilih memberikan hasil segmentasi terbaik atau tidak. Prinsip utama dari metode *Silhouette Score* adalah bahwa nilai skor yang lebih tinggi menunjukkan kualitas klusterisasi yang lebih baik, karena menunjukkan bahwa data dalam satu kluster semakin sama satu sama lain dan semakin jauh dari yang lain.



Gambar 14. Evaluasi *Silhouette Score*

Peneliti melakukan analisis terhadap berbagai jenis *cluster*, mulai dari dua hingga lima. Hasil analisis menunjukkan bahwa kelompok dengan 4 *cluster* memiliki nilai *Silhouette Score* tertinggi, sebesar 0,438, dibandingkan dengan kelompok lain. Hasil evaluasi ini menunjukkan bahwa pembagian data ke dalam empat *cluster* menghasilkan segmentasi pelanggan yang lebih terpisah dan keseragaman yang lebih tinggi dalam setiap *cluster*. Berdasarkan hasil ini, konfigurasi yang paling ideal untuk data yang dianalisis adalah pembagian *cluster* menjadi empat kelompok. Gambar 14 diatas menunjukkan detail hasil evaluasi menggunakan metode *Silhouette Score*.

4. CONCLUSION

Untuk memisahkan pelanggan berdasarkan fitur RFM, algoritma *clustering K-Means* berhasil digunakan. Dari 111 data, pembagian data menjadi 4 *cluster* dengan metode *Silhouette Score* memberikan hasil terbaik. *Cluster 4* terdiri dari pelanggan yang sangat loyal dengan nilai *Monetary* dan *Frequency* tinggi, sedangkan *Cluster 3* berisi pelanggan dengan nilai *Monetary* dan *Frequency* lebih rendah. *Cluster 2* mencakup pelanggan loyal dengan karakteristik mirip *Cluster 4*, dan *Cluster 1* terdiri dari pelanggan dengan potensi lebih rendah. Hasil *clustering* ini membantu manajemen merancang strategi pemasaran dan pengembangan bisnis. Penelitian ini memiliki beberapa keterbatasan, seperti penggunaan data yang hanya mencakup periode Januari hingga Oktober 2024 dan ketergantungan *K-Means* pada pemilihan jumlah *cluster*. Untuk penelitian selanjutnya, dapat dipertimbangkan untuk menggunakan algoritma *clustering* lain serta memperpanjang periode data, termasuk faktor eksternal seperti promosi atau musim, guna memperkaya analisis segmentasi pelanggan.

5. ACKNOWLEDGMENTS

Penulis dengan hormat mengucapkan terima kasih kepada semua pihak yang mendukung penelitian ini. Khususnya kepada manajemen PT XYZ atas izin dan kerja sama yang diberikan dalam penyediaan data penjualan yang menjadi dasar penelitian ini. Penulis juga mengucapkan terima kasih kepada tim administratif yang telah membantu dengan fasilitas dan sumber daya yang diperlukan selama penelitian, yang mempermudah pelaksanaan penelitian ini. Penulis juga mengucapkan terima kasih kepada dosen pembimbing yang telah memberikan bimbingan, arahan, dan masukan yang sangat penting, serta kepada rekan sejawat yang telah memberikan dukungan moral dan profesional selama proses penyusunan jurnal ini. Semua bentuk kontribusi yang diberikan sangat berharga bagi penyelesaian penelitian ini.

6. REFERENCES

- Adiana, B. E., Soesanti, I., & Permanasari, A. E. (2018). Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering. *JUTEI*, 2(1), 23–32. <https://doi.org/10.21460/jutei.2017.21.76>
- Auliasari, K., & Kertaningtyas, M. (2019). Penerapan Algoritma K-Means Untuk Segmentasi Konsumen Menggunakan R. *Jurnal Teknologi & Manajemen Informatika*, 5(1), 42–49. <https://doi.org/10.59134/jsk.v5i2.386>

- Awalina, E. F. L., & Rahayu, W. I. (2023). Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail. *Jurnal Teknologi Dan Informasi*, 13(2), 122–137. <https://doi.org/10.34010/jati.v13i2.10090>
- Bykova, E., Vasiliev, I., Bosneaga, V., & Suslov, V. (2024). Application of Neural Networks for Forecasting Energy Security Indicators. *Internasional Journal of Scientific Engineering and Research*, 12(11), 5–11.
- Fawzi, M. G. H., Iskandar, A. S., Erlangga, H., Nurjaya, & Denok, S. (2021). *Strategi Pemasaran*. Pascal Book.
- Gustriansyah, R., Suhandi, N., & Antony, F. (2020). Clustering optimization in RFM Analysis Based on K-Means. *Indonesian Journal of Electrical Engineering and Computer Science*, 18(1), 470–477. <https://doi.org/10.11591/ijeecs.v18.i1.pp470-477>
- Hananto, V. R., Churniawan, A. D., & Wardhanie, A. P. (2017). Perancangan Analytical CRM untuk Mendukung Segmentasi Pelanggan di Institusi Pendidikan. *Jurnal Ilmiah Teknologi Informasi Asia*, 11(1), 79–88. <https://doi.org/10.32815/jitika.v11i1.55>
- Hermawan, A., Jayanti, N. R., Saputra, A., Tambunan, C., Baihaqi, D. M., Syahreza, M. A., & Bachtiar, Z. (2024). Optimalisasi Strategi Pemasaran Melalui Analisis RFM pada Dataset Transaksi Ritel Menggunakan Python. *Jurnal Manajemen Riset Inovasi*, 2(4), 254–267.
- Kusumo, H., Sedyono, E., & Marwata, M. (2019). Analisis Algoritma Apriori Untuk Mendukung Strategi Promosi Perguruan Tinggi. *Walisono Journal of Information Technology*, 1(1), 51–62.
- Nursanti, T. D., Chatra, M. A., Adrian, Haitamy, A. G., Arisandi, D., Masdiantini, P. R., Waty, E., Boari, Y., & Judijanto, L. (2024). *Entrepreneurship*. PT Sonpedia Publishing Indonesia.
- Osborne, J. W. (2013). *Best Practice in Data Cleaning*. SAGE Publications.
- Perdana, Satria, A., Florentin, Sara, F., & Santoso, A. (2022). Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift. *Sebatik*, 26(2), 446–457. <https://doi.org/10.46984/sebatik.v26i2.2134>
- Reyes, M. (2020). *Consumer Behavior and Marketing*. Intech Open.
- Rizki, M., Devrika, D., Lubis, F. S., Silvia, & Umam, I. H. (2019). Aplikasi Data Mining dalam Penentuan Layout Swalayan dengan Menggunakan Metode MBA. *Jurnal Hasil Penelitian Dan Karya Ilmiah*, 5(2), 130–138.
- Sabuncu, I., Turkan, E., & Polat, H. (2020). Customer Segmentation and Profiling with RFM. *Turkish Journal of Marketing*, 5(1), 22–36.
- Sastya, N. C., & Nugraha, I. (2023). Penerapan Metode CRISP-DM dalam Menganalisis Data untuk Menentukan Customer Behavior di MeatSolution. *Jurnal Pendidikan Dan Aplikasi Industri*, 10(2), 103–115. <https://doi.org/10.33592/unistek.v10i2.3079>
- Setiawan, Z., Zebus, Rony, Sandra, Y., Suprayitno, D., Hamid, Rahmad, S., Islami, V., & Marsyaf, A. (2024). *Buku Ajar Perilaku Konsumen*. PT. Sonpedia Publishing Indonesia.
- Shaliha, K. M., Angelyna, Nugraha, A. A., Wahisyam, M. H., & Sandi, T. K. (2021). Implementasi K-Means Clustering pada Online Retail berdasarkan Recency , Frequency , dan Monetary. *Gunung Djati Conference Series*, 3(1), 99–106.
- Taqwim, W. A., Setiawan, N. Y., & Bachtiar, Fitra, A. (2019). Analisis Segmentasi Pelanggan Dengan RFM Model Pada Pt . Arthamas Citra Mandiri Menggunakan Metode Fuzzy C-Means Clustering. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(2), 1986–1993.
- Trianasari, N., & Permadi, T. A. (2024). Analysis of Product Recommendation Models at Each Fixed Broadband Sales Location Using K-Means, DBSCAN, Hierarchical Clustering, SVM, RF, and ANN. *Journal of Applied Data Sciences*, 5(2), 636–652. <https://doi.org/10.47738/jads.v5i2.210>
- Wahyuni, S., Wulansari, T. T., & Fahrullah, F. (2023). Segmentasi Pelanggan Berdasarkan Analisis Recency, Frequency, Monetary Menggunakan Algoritma K-Means Pada CV. Toedjoe Sinar Group. *Jurnal Rekayasa Teknologi Informasi (JURTI)*, 7(2), 180–187. <https://doi.org/10.30872/jurti.v7i2.8748>
- Wildemuth, B. M. (2017). *Application of Social Reserch Methods to Questions in Information and Library Science*. ABC-CLIO, LLC.
- Yunus, E. (2016). *Manajemen Strategis*. CV ANDI OFFSET.