

## Segmentasi Pelanggan dan Optimalisasi Penjualan pada Data Retail *Online* Berbasis Model *RFM*

Gunawan Budi Sulisty<sup>1</sup>, Noor Hasan<sup>2</sup>, Sri Kiswati<sup>\*3</sup>, Fransisca Natalia<sup>4</sup>, Elly Muningsih<sup>5</sup>

<sup>1,2,3,4,5</sup> Universitas Bina Sarana Informatika

Email: <sup>1</sup>gunawan.gnw@bsi.ac.id, <sup>2</sup>noor.nhs@bsi.ac.id, <sup>3</sup>sri.srk@bsi.ac.id, <sup>4</sup>fransisca.fia@bsi.ac.id, <sup>5</sup>elly.emh@bsi.ac.id

\*Penulis Korespondensi

### Abstrak

Penelitian ini dilatarbelakangi oleh pentingnya segmentasi pelanggan dalam meningkatkan efektivitas strategi pemasaran dan optimalisasi penjualan pada bisnis retail *online* yang semakin kompetitif. Segmentasi pelanggan memungkinkan perusahaan untuk mengelompokkan konsumen berdasarkan perilaku pembelian sehingga dapat merancang pendekatan pemasaran yang lebih tepat sasaran. Metode yang digunakan dalam penelitian ini adalah model *RFM* (*Recency, Frequency, Monetary*) yang mengukur nilai pelanggan berdasarkan waktu transaksi terakhir, frekuensi pembelian, dan nilai pembelian total. Data yang digunakan berasal dari dataset *Online Retail.csv* yang kemudian diproses dan dianalisis menggunakan algoritma *K-means clustering* untuk mengelompokkan pelanggan ke dalam segmen-segmen yang homogen. Hasil penelitian menunjukkan bahwa jumlah *cluster* optimal yang diperoleh adalah dua, dengan karakteristik pelanggan pada *Cluster 0* memiliki nilai *Recency* rendah dan *Frequency* tinggi, serta nilai *Monetary* yang lebih besar, sedangkan *Cluster 1* terdiri dari pelanggan dengan *Recency* tinggi dan *Frequency* rendah. Segmentasi ini memberikan gambaran yang jelas mengenai perilaku pelanggan yang berbeda dan memungkinkan perusahaan untuk mengarahkan strategi pemasaran yang lebih efektif. Penerapan model *RFM* yang dikombinasikan dengan *K-means clustering* terbukti efektif dalam segmentasi pelanggan retail *online*, sehingga dapat mendukung optimalisasi penjualan dan peningkatan loyalitas pelanggan. Temuan ini memiliki implikasi penting bagi pengembangan strategi pemasaran berbasis data yang lebih terarah dan efisien di industri retail *online*.

**Kata kunci:** *model rfm, segmentasi pelanggan, data retail online, k-means, optimalisasi penjualan, data mining*

### Abstract

*This study is motivated by the importance of customer segmentation in enhancing the effectiveness of marketing strategies and sales optimization in the increasingly competitive online retail business. Customer segmentation enables companies to group consumers based on purchasing behavior, allowing for more targeted marketing approaches. The method used in this research is the RFM (Recency, Frequency, Monetary) model, which measures customer value based on the time of the last transaction, purchase frequency, and total purchase value. The data used is derived from the Online Retail.csv dataset, which is processed and analyzed using the K-means clustering algorithm to group customers into homogeneous segments. The results show that the optimal number of clusters obtained is two, with customers in Cluster 0 characterized by low Recency and high Frequency, as well as higher Monetary value, while Cluster 1 consists of customers with high Recency and low Frequency. This segmentation provides a clear picture of different customer behaviors and enables companies to direct more effective marketing strategies. In conclusion, the application of the RFM model combined with K-means clustering proves effective in segmenting online retail customers, thereby supporting sales optimization and increasing customer loyalty. These findings have important implications for the development of more targeted and efficient data-driven marketing strategies in the online retail industry.*

**Keywords:** *rfm model, customer segmentation, online retail data, k-means, sales optimization, data mining*

## 1. PENDAHULUAN

Perkembangan pesat teknologi informasi dan komunikasi telah membawa perubahan signifikan dalam berbagai sektor, salah satunya adalah industri perdagangan, khususnya *e-commerce*. Di tingkat global, *e-commerce* telah menjadi salah satu pendorong utama transformasi digital yang mengubah cara konsumen berinteraksi dengan produk dan layanan. Di Indonesia, pertumbuhan *e-commerce* juga menunjukkan tren yang positif, didukung oleh penetrasi internet yang semakin luas, peningkatan penggunaan perangkat mobile, serta perubahan perilaku konsumen yang semakin mengutamakan kemudahan dan kecepatan dalam bertransaksi secara *online*. Fenomena ini tidak hanya meningkatkan volume transaksi secara *eksponensial*, tetapi juga memperkaya diversifikasi produk yang tersedia di *platform retail online*, mulai dari kebutuhan sehari-hari hingga produk-produk khusus dan *niche market*. Transformasi digital ini menghasilkan data transaksi dalam jumlah besar dan beragam, yang menjadi aset penting bagi pelaku bisnis untuk memahami perilaku konsumen dan mengoptimalkan strategi pemasaran serta penjualan (Chen et al., 2012).

Namun, di balik peluang besar tersebut, terdapat tantangan yang kompleks, terutama dalam hal segmentasi pelanggan. Platform retail *online* menghadapi keragaman perilaku konsumen yang sangat luas, mulai dari frekuensi pembelian, nilai transaksi, hingga waktu terakhir konsumen melakukan pembelian. Kompleksitas ini menuntut pendekatan segmentasi yang lebih canggih dan berbasis data untuk mengelompokkan pelanggan secara efektif. Segmentasi yang tepat sangat penting untuk meningkatkan efisiensi pemasaran, karena memungkinkan perusahaan untuk menyesuaikan strategi komunikasi dan penawaran produk sesuai dengan karakteristik dan kebutuhan masing-masing segmen pelanggan. Tanpa segmentasi yang akurat, upaya pemasaran dapat menjadi tidak efektif dan berpotensi membuang sumber daya yang berharga (Dibb, 1998).

Dalam konteks ini, data mining muncul sebagai alat strategis yang sangat berperan dalam bisnis ritel *online*. Data mining adalah proses eksplorasi dan analisis data dalam jumlah besar untuk menemukan pola, tren, dan hubungan yang tersembunyi yang tidak mudah terlihat melalui metode analisis konvensional. Dengan menggunakan teknik data mining, perusahaan dapat menggali wawasan yang lebih mendalam dari data transaksi yang kompleks dan beragam, sehingga mendukung pengambilan keputusan yang lebih tepat. Salah satu pendekatan yang populer dan efektif dalam segmentasi pelanggan adalah model RFM (Recency, Frequency, Monetary). Model ini mengukur nilai pelanggan berdasarkan tiga dimensi utama yakni seberapa baru pelanggan melakukan pembelian (*Recency*), seberapa sering pelanggan melakukan pembelian (*Frequency*), dan berapa besar nilai total pembelian pelanggan (*Monetary*). Pendekatan RFM memiliki keunggulan dibandingkan metode segmentasi tradisional karena lebih fokus pada perilaku aktual pelanggan dan memberikan gambaran yang lebih dinamis mengenai nilai pelanggan bagi perusahaan (Pine & Korn, 2011). Model RFM memungkinkan segmentasi yang lebih responsif terhadap perubahan perilaku konsumen dan lebih relevan untuk strategi pemasaran yang berbasis data.

Tujuan utama penelitian ini adalah mengaplikasikan model RFM (*Recency, Frequency, Monetary*) sebagai metode segmentasi pelanggan yang efektif pada dataset retail *online*, khususnya data transaksi yang terdapat dalam file *Online Retail.csv*. Model RFM dipilih karena kemampuannya dalam mengelompokkan pelanggan berdasarkan tiga dimensi utama yang mencerminkan perilaku pembelian mereka: *recency* (seberapa baru pelanggan melakukan pembelian terakhir), *frequency* (seberapa sering pelanggan melakukan pembelian dalam periode tertentu), dan *monetary* (berapa besar nilai total pembelian pelanggan). Dengan menggunakan model ini, penelitian bertujuan untuk mengidentifikasi pola perilaku konsumen yang tersembunyi dalam data transaksi, sehingga dapat mengelompokkan pelanggan ke dalam segmen-segmen yang memiliki karakteristik dan nilai bisnis yang berbeda-beda. Pendekatan ini memungkinkan pemahaman yang lebih mendalam mengenai perilaku pelanggan dan potensi kontribusi mereka terhadap pendapatan perusahaan, yang menjadi dasar penting dalam pengambilan keputusan pemasaran yang lebih terarah dan efisien (Susilawati Sugiana & Musty, 2023).

Selanjutnya, penelitian ini bertujuan untuk menginterpretasi hasil segmentasi tersebut guna mengidentifikasi segmen pelanggan bernilai tinggi yang dapat menjadi fokus utama dalam strategi pemasaran dan pengembangan produk. Dengan memahami karakteristik segmen-segmen ini, perusahaan dapat merancang program loyalitas, promosi, dan komunikasi yang lebih personal dan relevan, sehingga meningkatkan efektivitas interaksi dengan pelanggan dan memperkuat hubungan jangka panjang. Interpretasi hasil segmentasi juga akan membantu dalam mengidentifikasi segmen pelanggan yang berpotensi churn atau kurang aktif, sehingga dapat dilakukan intervensi yang tepat untuk mempertahankan mereka (Hendriyani, 2025). Selain itu, penelitian ini bertujuan memberikan rekomendasi optimalisasi

penjualan dan strategi pemasaran berdasarkan hasil segmentasi pelanggan. Rekomendasi ini mencakup strategi peningkatan nilai pelanggan melalui pendekatan yang disesuaikan dengan karakteristik masing-masing segmen, seperti penawaran produk khusus, diskon, atau program loyalitas yang dirancang untuk meningkatkan frekuensi pembelian dan nilai transaksi.

### 1.1. Ruang Lingkup Penelitian

Penelitian ini menggunakan *dataset Online Retail.csv* sebagai basis analisis utama untuk segmentasi pelanggan dan optimalisasi penjualan. Dataset ini merupakan kumpulan data transaksi retail *online* yang mencakup berbagai atribut penting, antara lain *InvoiceNo* (nomor faktur), *StockCode* (kode produk), *Description* (deskripsi produk), *Quantity* (jumlah barang yang dibeli), *InvoiceDate* (tanggal dan waktu transaksi), *UnitPrice* (harga per unit), *CustomerID* (identifikasi pelanggan), dan *Country* (negara asal pelanggan). Data ini mencakup periode waktu tertentu yang merekam transaksi historis secara rinci, memberikan gambaran komprehensif mengenai perilaku pembelian pelanggan di platform retail *online*. Cakupan geografis data ini juga cukup luas, meliputi berbagai negara, sehingga memungkinkan analisis yang lebih holistik terhadap pola pembelian lintas wilayah. Namun, perlu dicatat bahwa dataset ini memiliki keterbatasan, seperti tidak tersedianya data demografis pelanggan (misalnya usia, jenis kelamin, atau pendapatan) maupun variabel eksternal lain yang dapat memengaruhi perilaku konsumen, seperti faktor sosial ekonomi atau tren pasar saat itu. Keterbatasan ini menjadi pertimbangan penting dalam interpretasi hasil penelitian agar tidak berlebihan dalam generalisasi temuan (Kumar & Reinartz, 2016).

Fokus utama penelitian ini adalah penerapan teknik data mining dengan pendekatan model *RFM* (*Recency, Frequency, Monetary*) untuk melakukan segmentasi pelanggan dan analisis perilaku konsumen berdasarkan data transaksi historis yang tersedia. Model *RFM* dipilih karena kemampuannya dalam mengelompokkan pelanggan secara efektif berdasarkan tiga dimensi utama yang mencerminkan nilai dan loyalitas pelanggan terhadap bisnis. Penelitian ini secara khusus menitikberatkan pada implementasi dan evaluasi model *RFM* sebagai alat segmentasi yang sederhana namun *powerful*. Dengan demikian, ruang lingkup penelitian dibatasi pada analisis data transaksi historis tanpa memasukkan variabel eksternal atau data tambahan yang dapat memperkaya model, sehingga hasil yang diperoleh lebih fokus dan spesifik pada perilaku pembelian yang tercermin dalam dataset *Online Retail.csv* (Widyaningsih & Rusdiana, 2025).

## 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksplanatori-eksploratif untuk menganalisis perilaku pelanggan berdasarkan data transaksi *online* retail. Model *RFM* digunakan untuk membentuk fitur pelanggan, kemudian dilakukan segmentasi menggunakan algoritma *K-Means Clustering*. Tahapan penelitian meliputi *preprocessing data*, perhitungan skor *RFM*, normalisasi, penentuan jumlah kluster optimal, dan penerapan *K-Means*. Selain itu, algoritma *Decision Tree* digunakan untuk memprediksi potensi churn pelanggan. Seluruh proses dianalisis dan divisualisasikan menggunakan *Python*.

### 2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksplanatori-eksploratif yang bertujuan untuk menemukan pola tersembunyi dalam perilaku pelanggan berdasarkan data transaksi *online retail*. Analisis dilakukan melalui proses segmentasi pelanggan menggunakan teknik *unsupervised machine learning*, khususnya algoritma *K-Means Clustering*. Pendekatan ini dipilih karena mampu mengelompokkan data berdimensi banyak ke dalam kluster yang representatif tanpa memerlukan label kelas sebagai acuan (Awalina & Rahayu, 2023).

Penelitian ini dilaksanakan melalui tahapan berikut: (1) pengumpulan dan pra-pemrosesan data, (2) pembuatan fitur menggunakan metode *Recency, Frequency, Monetary (RFM)*, (3) normalisasi data, (4) penentuan jumlah kluster optimal menggunakan *Elbow Method* dan *Silhouette Score*, (5) penerapan algoritma *K-Means*, dan (6) interpretasi hasil segmentasi.



Gambar 1. Tahapan penelitian

Model *RFM* merupakan metode yang banyak digunakan dalam analisis perilaku pelanggan karena kemampuannya dalam mengelompokkan pelanggan berdasarkan tiga dimensi utama yang mencerminkan nilai dan loyalitas pelanggan (Wafda, 2024). Dimensi *Recency* (*R*) mengukur seberapa baru pelanggan melakukan transaksi terakhir, *Frequency* (*F*) mengukur seberapa sering pelanggan melakukan pembelian dalam periode tertentu, dan *Monetary* (*M*) mengukur total nilai pembelian pelanggan selama periode tersebut. Dengan menggunakan model ini, perusahaan dapat mengidentifikasi kelompok pelanggan yang berbeda dan merancang strategi pemasaran yang lebih terfokus dan efektif. Secara matematis, variabel *RFM* dirumuskan sebagai berikut:

*Recency* (*R*) dihitung sebagai selisih antara tanggal maksimum transaksi dalam dataset dengan tanggal transaksi terakhir pelanggan, yaitu  $R=T-t_i$  di mana (*T*) adalah tanggal transaksi terbaru dalam dataset dan (*t<sub>i</sub>*) adalah tanggal transaksi terakhir pelanggan.

*Frequency* (*F*) adalah jumlah transaksi unik yang dilakukan oleh pelanggan dalam periode pengamatan. Rumus menghitung *F* dapat dilihat pada persamaan (1).

$$F = \sum_{j=1}^n x_j \quad (1)$$

*Monetary* (*M*) merupakan total nilai pembelian pelanggan yang dihitung sebagai jumlah dari hasil perkalian kuantitas barang dengan harga satuan pada setiap transaksi, dimana (*q<sub>k</sub>*) adalah kuantitas item ke- (*k*), dan (*p<sub>k</sub>*) adalah harga unit item ke- (*k*). Rumus menghitung *Monetary* dapat dilihat pada persamaan (2).

$$M = \sum_{k=1}^m (q_k \times p_k) \quad (2)$$

## 2.2. Preprocessing Data

Tahapan awal dalam metode penelitian ini adalah *preprocessing* data yang bertujuan memastikan kualitas dan konsistensi data sebelum diterapkan model *RFM* dan teknik segmentasi pelanggan. Dataset transaksi retail *online* yang digunakan mengandung berbagai masalah data seperti *missing value*, *outlier*, dan variasi skala data yang memengaruhi hasil analisis. Oleh karena itu, proses *preprocessing* menjadi langkah krusial untuk meningkatkan akurasi dan reliabilitas model yang akan dibangun.

Penanganan *missing value* merupakan langkah pertama yang harus dilakukan. *Missing value* dapat muncul akibat kesalahan pencatatan, data yang hilang, atau ketidaksesuaian format (Dawis et al., 2022). Dalam penelitian ini, *missing value* pada kolom penting seperti *CustomerID*, *Quantity*, *UnitPrice*, dan *InvoiceDate* akan diidentifikasi dan ditangani. Pendekatan yang umum digunakan adalah menghapus baris data yang memiliki *missing value* pada kolom-kolom kritis tersebut, karena imputasi data pada variabel transaksi sering kali berisiko menimbulkan bias. Namun, jika *missing value* ditemukan dalam jumlah kecil dan tidak signifikan, penghapusan data ini tidak akan mengurangi representativitas dataset secara keseluruhan.

Setelah *missing value* ditangani, langkah berikutnya adalah deteksi dan penanganan *outlier*. *Outlier* dalam data transaksi dapat berupa nilai *Quantity* atau *UnitPrice* yang sangat ekstrem, baik sangat besar maupun negatif (misalnya transaksi pembatalan atau retur). *Outlier* ini dapat mengganggu perhitungan skor *RFM* dan mempengaruhi segmentasi pelanggan. Metode yang digunakan untuk mendeteksi *outlier* adalah *Interquartile Range (IQR)*, yaitu dengan menghitung rentang antar kuartil (*Q3 - Q1*) dan menetapkan batas bawah dan batas atas.

Setelah data bersih dari *missing value* dan *outlier*, tahap selanjutnya adalah normalisasi data. Normalisasi diperlukan karena variabel *R*, *F*, dan *M* memiliki skala yang berbeda dan rentang nilai yang bervariasi. Untuk menghindari dominasi variabel dengan skala besar dalam proses *clustering* dan *scoring*, data dinormalisasi menggunakan metode *Min-Max Scaling* yang mengubah nilai variabel ke rentang 0.

## 2.3. Penerapan Model RFM dan Klasifikasi Pelanggan

Model *RFM* (*Recency*, *Frequency*, *Monetary*) merupakan metode yang sangat efektif untuk mengukur nilai dan perilaku pelanggan berdasarkan data transaksi (Wicaksana et al., 2022). Dalam penelitian ini, penerapan model *RFM* dilakukan dengan mengkalkulasi tiga dimensi utama yang merepresentasikan karakteristik pelanggan secara kuantitatif. Setiap dimensi kemudian diklasifikasikan ke dalam skor yang merefleksikan tingkat nilai pelanggan pada aspek tersebut, dengan skala 1 sampai 5, di mana skor 5 menunjukkan nilai tertinggi dan 1 nilai terendah.

Dimensi pertama, *Recency* ( $R$ ), mengukur seberapa baru pelanggan melakukan transaksi terakhirnya. Secara matematis, *Recency* dihitung sebagai selisih antara tanggal transaksi terbaru dalam dataset dengan tanggal transaksi terakhir pelanggan, karena mereka dianggap lebih aktif dan berpotensi lebih berharga.

Dimensi kedua, *Frequency* ( $F$ ), mengukur seberapa sering pelanggan melakukan pembelian dalam periode pengamatan. *Frequency* dihitung sebagai jumlah transaksi unik yang dilakukan pelanggan dengan frekuensi pembelian tinggi akan mendapatkan skor  $F$  yang tinggi, menandakan loyalitas dan keterlibatan yang kuat dengan toko *online*.

Dimensi ketiga, *Monetary* ( $M$ ), mengukur total nilai pembelian pelanggan selama periode tertentu. Nilai ini dihitung dengan menjumlahkan hasil perkalian kuantitas barang dengan harga satuan pada setiap transaksi pelanggan. Pelanggan dengan nilai pembelian total yang besar akan memperoleh skor  $M$  yang tinggi, menunjukkan kontribusi finansial yang signifikan.

Setelah nilai  $R$ ,  $F$ , dan  $M$  dihitung untuk setiap pelanggan, langkah berikutnya adalah mengklasifikasikan nilai-nilai tersebut ke dalam lima level skor (1 sampai 5). Pembagian skor ini dilakukan dengan membagi rentang nilai masing-masing dimensi menjadi lima interval yang mewakili tingkat rendah hingga tinggi. Misalnya, untuk *Recency*, pelanggan dengan nilai selisih hari paling kecil (transaksi terbaru) akan mendapatkan skor 5, sedangkan yang paling lama mendapatkan skor 1. Demikian pula untuk *Frequency* dan *Monetary*, pelanggan dengan nilai tertinggi mendapatkan skor 5, dan sebaliknya.

Penerapan model RFM ini sangat penting karena memberikan gambaran yang jelas dan terukur mengenai perilaku pelanggan, memungkinkan perusahaan untuk mengelompokkan pelanggan ke dalam segmen-segmen yang bermakna. Dengan segmentasi yang tepat, strategi pemasaran dapat disesuaikan untuk meningkatkan retensi pelanggan, mengoptimalkan penawaran produk, dan pada akhirnya meningkatkan penjualan dan profitabilitas perusahaan.

Selain itu, skor RFM yang telah dihitung akan menjadi input utama untuk tahap segmentasi pelanggan menggunakan algoritma clustering, yang akan dijelaskan pada bagian selanjutnya. Dengan demikian, penerapan model RFM tidak hanya berfungsi sebagai alat pengukuran, tetapi juga sebagai fondasi analitis untuk pengelompokan pelanggan yang lebih lanjut dan pengembangan strategi bisnis yang lebih efektif.

#### 2.4. Segmentasi Pelanggan dengan Algoritma *K-means*

Setelah skor RFM dihitung dan pelanggan diklasifikasikan berdasarkan nilai *Recency*, *Frequency*, dan *Monetary*, langkah berikutnya adalah melakukan segmentasi pelanggan menggunakan algoritma *K-means clustering*. Metode ini dipilih karena kemampuannya dalam mengelompokkan data ke dalam beberapa kluster berdasarkan kemiripan fitur, sehingga dapat mengidentifikasi kelompok pelanggan dengan karakteristik perilaku pembelian yang serupa secara efektif dan efisien. Rumus untuk menghitung segmentasi pelanggan pada persamaan (3).

$$J = \sum_{l=1}^k \sum_{x \in C_l} |x - \mu_l|^2 \quad (3)$$

Proses segmentasi dimulai dengan menentukan jumlah kluster optimal yang akan digunakan dalam algoritma *K-means*. Penentuan jumlah kluster ini sangat penting karena jumlah kluster yang terlalu sedikit dapat mengaburkan perbedaan antar segmen pelanggan, sedangkan jumlah kluster yang terlalu banyak dapat menyebabkan overfitting dan kesulitan interpretasi. Untuk itu, metode *Elbow* digunakan sebagai teknik evaluasi untuk menemukan titik optimal jumlah kluster. Metode *Elbow* bekerja dengan cara menjalankan algoritma *K-means* untuk berbagai jumlah kluster (misalnya dari 1 hingga 10) dan menghitung nilai *Within-Cluster Sum of Squares* (WCSS) untuk setiap jumlah kluster. Grafik WCSS terhadap jumlah kluster kemudian dianalisis untuk menemukan titik "*elbow*" atau titik di mana penurunan WCSS mulai melambat secara signifikan. Titik ini dianggap sebagai jumlah kluster optimal yang seimbang antara kompleksitas model dan kualitas segmentasi.

Setelah jumlah kluster optimal ditentukan, algoritma *K-means* dijalankan pada data skor RFM yang telah dinormalisasi. Algoritma ini bekerja dengan menginisialisasi *centroid* secara acak, kemudian mengelompokkan setiap pelanggan ke *centroid* terdekat berdasarkan jarak *Euclidean*. Proses iteratif ini berlanjut hingga *centroid* tidak mengalami perubahan signifikan atau mencapai batas iterasi maksimum. Hasilnya adalah pembagian pelanggan ke dalam beberapa kluster yang masing-masing memiliki karakteristik RFM yang homogen.

Setiap kluster yang terbentuk kemudian dianalisis untuk memahami karakteristik pelanggan di dalamnya. Misalnya, satu kluster mungkin terdiri dari pelanggan dengan skor *Recency* dan *Frequency* tinggi serta nilai *Monetary* besar, yang menunjukkan pelanggan loyal dan bernilai tinggi. Kluster lain mungkin berisi pelanggan dengan skor *Recency* rendah dan *Frequency* rendah, yang mengindikasikan pelanggan yang sudah lama tidak aktif dan berisiko *churn*. Analisis ini sangat penting untuk memberikan gambaran segmentasi yang jelas dan actionable bagi perusahaan.

Segmentasi pelanggan yang dihasilkan dari proses ini memungkinkan perusahaan untuk mengembangkan strategi pemasaran yang lebih terarah dan efektif. Misalnya, pelanggan di kluster bernilai tinggi dapat diberikan program loyalitas khusus atau penawaran eksklusif, sementara pelanggan di kluster berisiko *churn* dapat diberikan insentif untuk mengaktifkan kembali pembelian mereka. Dengan demikian, segmentasi ini tidak hanya meningkatkan pemahaman terhadap perilaku pelanggan, tetapi juga mendukung pengambilan keputusan bisnis yang lebih tepat sasaran.

Optimalisasi penjualan merupakan tahap krusial yang memanfaatkan hasil segmentasi pelanggan untuk merancang strategi pemasaran yang lebih efektif dan efisien (Apriana & Yuliansyah, 2024). Setelah pelanggan dikelompokkan berdasarkan skor RFM dan kluster yang terbentuk melalui algoritma K-means, setiap segmen dianalisis karakteristiknya secara mendalam untuk menentukan pendekatan pemasaran yang sesuai. Misalnya, segmen pelanggan dengan skor RFM tinggi yang menunjukkan loyalitas dan nilai pembelian besar dapat diberikan program loyalitas eksklusif, diskon khusus, atau penawaran produk premium untuk mempertahankan dan meningkatkan keterikatan mereka. Sebaliknya, segmen dengan skor rendah atau pelanggan yang berpotensi *churn* dapat diberikan insentif seperti kupon diskon, kampanye reaktivasi, atau komunikasi personal untuk mengurangi risiko kehilangan pelanggan.

Selain strategi pemasaran berbasis segmentasi, penelitian ini juga mengintegrasikan model prediksi *churn rate* menggunakan algoritma *Decision Tree*. Prediksi *churn* bertujuan mengidentifikasi pelanggan yang berisiko berhenti bertransaksi sehingga perusahaan dapat mengambil tindakan preventif lebih awal. Algoritma *Decision Tree* dipilih karena kemampuannya dalam menghasilkan model yang mudah diinterpretasikan dan efektif dalam klasifikasi. Data input untuk model ini meliputi variabel RFM, frekuensi interaksi, dan riwayat pembelian pelanggan. Model ini dilatih dan diuji menggunakan data historis, kemudian dievaluasi dengan metrik seperti *Accuracy* dan *Precision* untuk memastikan keandalan prediksi.

Evaluasi model prediksi *churn* sangat penting agar perusahaan dapat mempercayai hasilnya dalam pengambilan keputusan. *Accuracy* mengukur proporsi prediksi yang benar dari keseluruhan data, sedangkan *Precision* menilai ketepatan prediksi positif *churn*, yaitu seberapa banyak pelanggan yang diprediksi akan *churn* benar-benar berhenti bertransaksi. Dengan metrik evaluasi ini, model dapat dioptimalkan untuk meminimalkan kesalahan prediksi yang berpotensi merugikan perusahaan.

Integrasi antara segmentasi pelanggan dan prediksi *churn* memberikan pendekatan holistik dalam optimalisasi penjualan. Segmentasi memungkinkan perusahaan memahami karakteristik dan kebutuhan setiap kelompok pelanggan, sementara prediksi *churn* membantu mengidentifikasi risiko kehilangan pelanggan secara individual. Dengan demikian, perusahaan dapat merancang strategi pemasaran yang tidak hanya meningkatkan penjualan tetapi juga mempertahankan loyalitas pelanggan secara berkelanjutan. Pendekatan ini juga memberikan nilai tambah berupa kemampuan adaptasi terhadap perubahan perilaku pelanggan dan dinamika pasar yang cepat, sehingga perusahaan retail *online* dapat tetap kompetitif dan responsif terhadap kebutuhan pelanggan (Yusepa et al., 2024).

## 2.5. Desain Sistem dan Arsitektur Implementasi

Untuk mengimplementasikan metode penelitian yang menggabungkan model RFM dan teknik segmentasi pelanggan berbasis clustering serta prediksi *churn*, diperlukan sebuah desain sistem yang terstruktur dan terintegrasi secara menyeluruh. Desain sistem ini bertujuan untuk memfasilitasi alur kerja mulai dari pengumpulan data, *preprocessing*, *scoring* RFM, segmentasi pelanggan, hingga visualisasi hasil dan rekomendasi strategi pemasaran yang dapat dioperasikan secara praktis dalam lingkungan bisnis retail *online*.

Arsitektur sistem yang diusulkan terdiri dari beberapa komponen utama yang saling berkesinambungan. Pertama, tahap pengumpulan dan ekstraksi data dilakukan dengan mengambil data transaksi mentah dari sumber dataset penjualan *Online Retail.csv*. Data ini kemudian masuk ke modul *preprocessing* yang bertugas membersihkan data, menangani *missing value*, mendeteksi dan mengeliminasi *outlier*, serta melakukan normalisasi agar data siap untuk dianalisis.

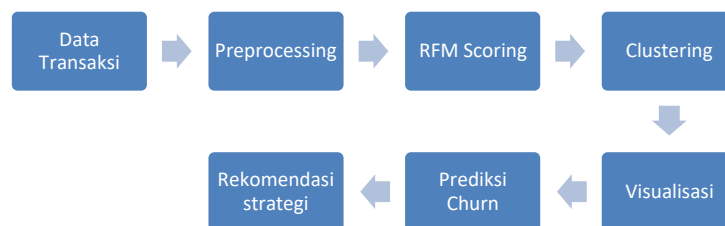
Setelah data siap, modul *scoring RFM* menghitung nilai *Recency*, *Frequency*, dan *Monetary* untuk setiap pelanggan berdasarkan rumus matematis yang telah ditetapkan. Nilai-nilai ini kemudian diklasifikasikan ke dalam skor 1 sampai 5 untuk masing-masing dimensi, membentuk matriks RFM yang

menjadi dasar segmentasi pelanggan. Modul segmentasi selanjutnya menggunakan algoritma *K-means clustering* untuk mengelompokkan pelanggan ke dalam beberapa kluster berdasarkan skor RFM yang telah dinormalisasi.

Hasil segmentasi kemudian divisualisasikan menggunakan grafik dan dashboard interaktif yang memudahkan pemahaman pola pelanggan dan karakteristik setiap kluster. Visualisasi ini berupa *scatter plot*, dan diagram batang yang menampilkan distribusi skor RFM dan ukuran kluster. Selain itu, modul prediksi *churn* menggunakan algoritma *Decision Tree* diintegrasikan untuk mengidentifikasi pelanggan yang berisiko berhenti bertransaksi, sehingga perusahaan dapat mengambil tindakan preventif yang tepat sasaran.

Tahap akhir dari sistem adalah modul rekomendasi strategi pemasaran yang memberikan saran berbasis data untuk setiap segmen pelanggan, berupa penawaran diskon khusus, program loyalitas, atau iklan reaktivasi. Modul ini dapat dihubungkan dengan sistem CRM atau platform pemasaran digital untuk otomatisasi pelaksanaan strategi.

Teknologi yang digunakan dalam implementasi sistem ini meliputi bahasa pemrograman *Python* dengan pustaka-pustaka utama seperti *Pandas* untuk manipulasi data, *Scikit-learn* untuk algoritma machine learning (*K-means*, *Decision Tree*), dan *Matplotlib* untuk visualisasi data. Penggunaan teknologi *open-source* ini memungkinkan fleksibilitas, skalabilitas, dan kemudahan replikasi penelitian oleh peneliti lain maupun praktisi bisnis. Secara keseluruhan, arsitektur sistem ini dirancang agar modular dan mudah dikembangkan, memungkinkan integrasi dengan sumber data lain atau penambahan metode analisis lanjutan di masa depan. Diagram arsitektur sistem dapat digambarkan sebagai alur berikut:

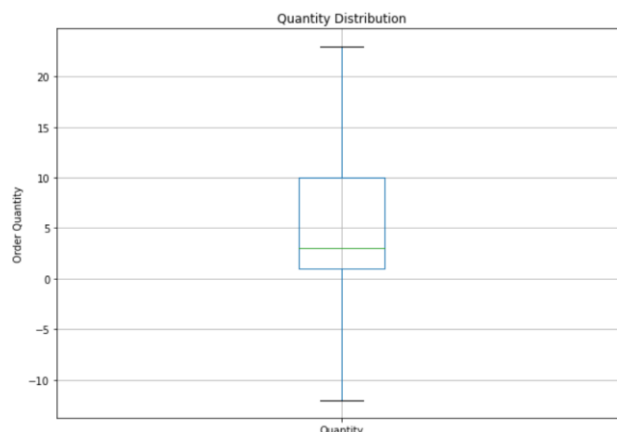


Gambar 2. Diagram arsitektur sistem

Dengan desain sistem yang komprehensif ini, penelitian tidak hanya memberikan kontribusi teoritis tetapi juga solusi praktis yang dapat diadopsi oleh perusahaan retail *online* untuk meningkatkan pemahaman pelanggan dan mengoptimalkan penjualan secara signifikan.

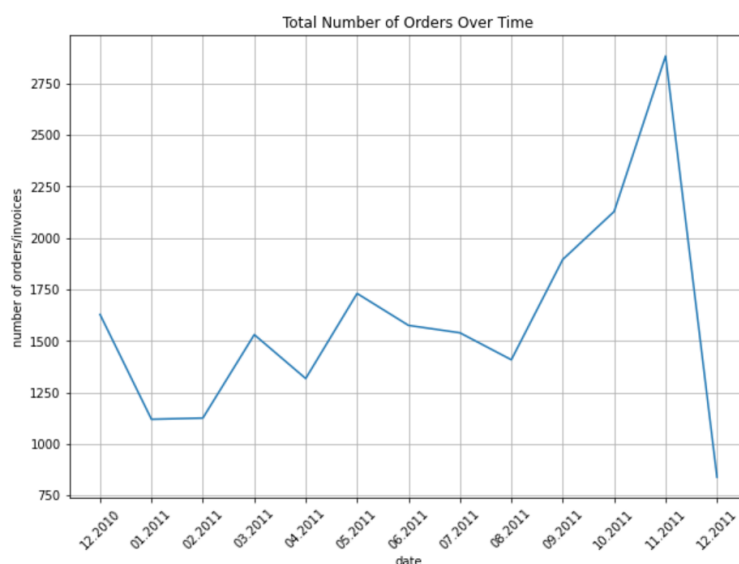
### 3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian kali ini menggunakan *Online Retail Dataset* dari *UCI Machine Learning Repository*, yang berisikan seluruh transaksi yang dilakukan menggunakan transaksi *online*. Untuk melaksanakan seleksi data kita menggunakan bahasa pemrograman *Python*. Seleksi data menggunakan *Python* bertujuan untuk mengambil data yang benar-benar relevan dan diperlukan dalam proses analisis atau pemodelan. Langkah ini penting untuk menyaring kolom (*fitur*) maupun baris (*record*) dari *dataset* yang memiliki kontribusi signifikan terhadap tujuan penelitian, seperti prediksi produk terlaris. Melalui seleksi data, informasi yang tidak lengkap, tidak valid, atau bersifat *noise* seperti nilai kosong, duplikat, dan nilai ekstrem (*outlier*) dapat dihilangkan, sehingga kualitas data meningkat dan akurasi model yang dibangun pun menjadi lebih baik. Selain itu, seleksi data juga memungkinkan peneliti untuk fokus pada *subset* data tertentu, misalnya transaksi dari negara atau periode waktu tertentu, yang sesuai dengan konteks studi. Dengan menyederhanakan dataset melalui seleksi yang tepat, proses pelatihan model menjadi lebih efisien, cepat, dan interpretatif, serta meminimalkan risiko *overfitting* akibat fitur yang tidak relevan.



Gambar 3. Grafik *Quantity Distribution*

Dari gambar 3. diatas banyaknya penjualan berada di antara *range* 0 sampai 10, namun dari data tersebut masih ada yang nilainya minus. Selanjutnya kita melihat deskripsi data, dari kolom *Quantity* terdapat sebanyak 541000 data. Untuk itu kita harus menghapus produk yang penjualannya minus dimana itu terjadi dikarenakan *refund*.



Gambar 4. Grafik *Time-series Number of Orders*

Penelitian ini berhasil melakukan segmentasi pelanggan pada data retail *online* menggunakan model *RFM* (*Recency, Frequency, Monetary*) yang dipadukan dengan algoritma *K-Means* clustering. Analisis *RFM* dilakukan dengan menghitung nilai *Recency, Frequency, dan Monetary* untuk setiap pelanggan, kemudian dinormalisasi ke skala 0-5 agar dapat digunakan sebagai input *clustering*. Pemilihan jumlah kluster optimal didasarkan pada nilai *Silhouette Score*, yang menunjukkan bahwa dua kluster memberikan segmentasi terbaik dengan nilai 0,7279, menandakan pemisahan kluster yang cukup jelas dan kohesif.

Tabel berikut menunjukkan centroid kluster yang merepresentasikan nilai rata-rata *Recency, Frequency, dan Monetary* pada masing-masing kluster:

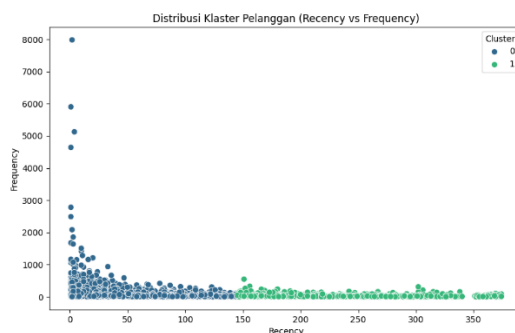
Tabel 1. *Centroid kluster*

<b>Cluster</b>	<b>Recency (Normalisasi)</b>	<b>Frequency (Normalisasi)</b>	<b>Monetary (Normalisasi)</b>
0	0.106	0.014	0.024
1	0.664	0.004	0.017

Dari tabel tersebut, terlihat bahwa *Cluster 0* memiliki nilai *Recency* yang sangat rendah, menunjukkan pelanggan yang baru saja melakukan transaksi, serta nilai *Frequency* dan *Monetary* yang relatif lebih tinggi dibandingkan *Cluster 1*. Sebaliknya, *Cluster 1* memiliki nilai *Recency* yang tinggi, menandakan pelanggan yang sudah lama tidak bertransaksi, dengan frekuensi dan nilai pembelian yang rendah.

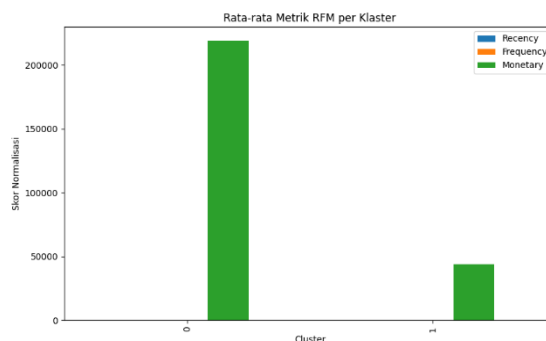
Evaluasi tambahan menggunakan metrik inerti menunjukkan nilai sebesar 72,58, yang mengindikasikan total jarak kuadrat dalam kluster relatif kecil, sehingga kluster yang terbentuk cukup kompak dan representatif.

Pada *scatter plot*, pelanggan dalam *Cluster 0* ditandai dengan warna biru, tersebar pada area dengan *Recency* rendah dan *Frequency* tinggi, sedangkan pelanggan dalam *Cluster 1* berwarna hijau, tersebar pada area dengan *Recency* tinggi dan *Frequency* rendah.



Gambar 5. *Scatter Plot* Distribusi Kalster Pelanggan

Berikut grafik batang yang membandingkan rata-rata nilai RFM per kluster memperlihatkan perbedaan signifikan pada nilai *Monetary*, di mana *Cluster 0* memiliki nilai pembelian rata-rata yang jauh lebih tinggi dibandingkan *Cluster 1*. Hal ini menegaskan bahwa pelanggan dalam *Cluster 0* merupakan segmen yang paling berharga bagi bisnis.



Gambar 6. Grafik batang Rata-rata metrik RFM per Kluster

Pelanggan dalam *Cluster 0* merupakan segmen loyal dengan aktivitas transaksi yang tinggi dan nilai pembelian besar, sesuai dengan teori segmentasi pelanggan yang menekankan pentingnya mempertahankan pelanggan bernilai tinggi untuk meningkatkan *lifetime value* (Adha, 2024). Sebaliknya, pelanggan dalam *Cluster 1* adalah segmen yang perlu diaktifkan kembali melalui strategi pemasaran khusus, seperti kampanye reaktivasi, sesuai dengan prinsip pemasaran relasional (Gustiasari, 2024).

Hasil segmentasi pelanggan berdasarkan model RFM dan algoritma *K-Means* menunjukkan adanya dua kluster utama dengan karakteristik yang berbeda secara signifikan. Kluster pertama (*Cluster 0*) memiliki nilai *Recency* yang rendah, *Frequency* yang tinggi, dan *Monetary* yang relatif besar. Hal ini mengindikasikan bahwa pelanggan dalam kluster ini adalah pelanggan yang baru saja melakukan pembelian, sering bertransaksi, dan memberikan kontribusi pendapatan yang tinggi. Karakteristik ini sesuai dengan konsep pelanggan loyal dan bernilai tinggi dalam teori segmentasi pelanggan, yang menjadi fokus utama dalam strategi pemasaran untuk mempertahankan dan meningkatkan nilai pelanggan (Haque et al., 2022). Pelanggan dengan profil seperti ini cenderung memiliki tingkat retensi yang tinggi dan potensi pembelian berulang yang signifikan, sehingga menjadi aset penting bagi perusahaan retail *online*.

Sebaliknya, klaster kedua (*Cluster 1*) memiliki nilai *Recency* yang tinggi, *Frequency* yang rendah, dan *Monetary* yang kecil. Ini menunjukkan pelanggan yang sudah lama tidak melakukan transaksi, jarang berbelanja, dan memberikan kontribusi pendapatan yang rendah. Pelanggan dalam segmen ini berpotensi mengalami *churn* atau kehilangan minat terhadap produk dan layanan yang ditawarkan. Oleh karena itu, mereka memerlukan pendekatan pemasaran yang berbeda, seperti kampanye reaktivasi, penawaran diskon khusus, atau komunikasi personalisasi untuk mengembalikan minat dan meningkatkan frekuensi pembelian. Pendekatan ini sejalan dengan prinsip pemasaran relasional yang menekankan pentingnya mempertahankan pelanggan lama dan mengurangi tingkat *churn* (Peppers & Rogers, 2011).

Interpretasi hasil ini juga menguatkan pentingnya penggunaan model RFM sebagai alat ukur perilaku pelanggan yang efektif dalam mengidentifikasi segmen pelanggan yang berbeda berdasarkan nilai dan aktivitas mereka. Dengan menggabungkan RFM dan K-Means, segmentasi menjadi lebih objektif dan berbasis data, memungkinkan perusahaan untuk mengelompokkan pelanggan secara otomatis dan akurat sesuai dengan pola transaksi mereka. Hal ini memberikan keunggulan dibandingkan metode segmentasi tradisional yang seringkali subjektif dan kurang terukur.

Selain itu, hasil segmentasi ini memberikan dasar yang kuat untuk pengambilan keputusan bisnis yang lebih tepat sasaran. Misalnya, perusahaan dapat mengalokasikan anggaran pemasaran secara efisien dengan memfokuskan program loyalitas dan penawaran eksklusif pada pelanggan bernilai tinggi (*Cluster 0*), sementara mengembangkan strategi reaktivasi dan komunikasi khusus untuk pelanggan kurang aktif (*Cluster 1*). Pendekatan ini tidak hanya meningkatkan efektivitas pemasaran tetapi juga berpotensi meningkatkan pendapatan dan retensi pelanggan secara keseluruhan.

Hasil segmentasi pelanggan yang diperoleh dari analisis RFM dan *algoritma K-Means* memberikan landasan yang kuat untuk merancang strategi pemasaran yang lebih efektif dan efisien dalam konteks retail *online*. Dengan memahami karakteristik dan perilaku masing-masing segmen pelanggan, perusahaan dapat mengalokasikan sumber daya pemasaran secara lebih tepat sasaran, sehingga meningkatkan efektivitas kampanye dan memaksimalkan *Return On Investment (ROI)*.

Pelanggan dalam klaster dengan nilai RFM tinggi, yang menunjukkan aktivitas transaksi yang baru dan sering serta nilai pembelian yang besar, merupakan aset paling berharga bagi perusahaan. Oleh karena itu, strategi pemasaran yang diarahkan kepada segmen ini harus fokus pada program loyalitas yang mampu mempertahankan dan meningkatkan keterikatan pelanggan. Contohnya adalah pemberian reward eksklusif, diskon khusus, akses awal ke produk baru, atau layanan pelanggan premium. Pendekatan ini tidak hanya meningkatkan kepuasan dan loyalitas pelanggan, tetapi juga mendorong peningkatan frekuensi pembelian dan nilai transaksi, yang secara langsung berkontribusi pada peningkatan pendapatan perusahaan.

Sebaliknya, pelanggan dalam klaster dengan nilai RFM rendah, yang menunjukkan jarak waktu transaksi yang lama, frekuensi rendah, dan nilai pembelian kecil, memerlukan pendekatan yang berbeda. Strategi reaktivasi menjadi sangat penting untuk mengembalikan minat dan keterlibatan pelanggan ini. Kampanye pemasaran yang dipersonalisasi, seperti penawaran diskon khusus, pengingat produk yang relevan, atau komunikasi yang menekankan manfaat produk, dapat digunakan untuk menarik kembali pelanggan yang mulai menghilang. Dengan demikian, perusahaan dapat mengurangi tingkat *churn* dan meningkatkan retensi pelanggan, yang merupakan faktor kunci dalam mempertahankan pendapatan jangka panjang.

Selain itu, segmentasi pelanggan memungkinkan perusahaan untuk mengoptimalkan alokasi anggaran pemasaran. Dengan mengetahui proporsi dan karakteristik tiap segmen, perusahaan dapat menentukan prioritas pengeluaran yang lebih rasional, misalnya mengalokasikan lebih banyak dana untuk program loyalitas pelanggan bernilai tinggi dan kampanye reaktivasi yang terukur untuk pelanggan kurang aktif. Pendekatan ini menghindari pemborosan sumber daya pada segmen yang kurang potensial dan meningkatkan efisiensi operasional pemasaran.

Lebih jauh, pendekatan berbasis data ini membuka peluang untuk pengembangan strategi pemasaran yang lebih dinamis dan responsif terhadap perubahan perilaku pelanggan. Dengan pemantauan berkala terhadap metrik RFM dan segmentasi ulang secara periodik, perusahaan dapat menyesuaikan strategi pemasaran secara real-time, sehingga mampu merespons tren pasar dan kebutuhan pelanggan dengan lebih cepat dan tepat.

#### 4. KESIMPULAN

Penelitian ini berhasil melakukan segmentasi pelanggan pada data retail *online* menggunakan model RFM yang dikombinasikan dengan algoritma *K-Means clustering*. Hasil segmentasi menunjukkan adanya

perbedaan karakteristik pelanggan yang signifikan berdasarkan metrik *Recency*, *Frequency*, dan *Monetary*, yang memungkinkan identifikasi segmen pelanggan dengan nilai tinggi maupun segmen yang memerlukan perhatian khusus untuk reaktivasi. Dengan demikian, model RFM terbukti efektif dalam mengelompokkan pelanggan secara meaningful untuk mendukung optimalisasi strategi penjualan. Implikasi dari penelitian ini adalah perusahaan retail *online* dapat menggunakan segmentasi tersebut untuk mengarahkan strategi pemasaran yang lebih tepat sasaran, meningkatkan loyalitas pelanggan, dan memaksimalkan pendapatan. Penelitian ini juga membuka peluang pengembangan model analisis pelanggan yang lebih kompleks dan prediktif di masa depan.

## REFERENSI

- Adha, E. I. D. (2024). *Implementasi Algoritma K-Means dengan Optimasi SOM Dalam Pemetaan Karakteristik Pelanggan Berdasarkan Model LRFM pada Bisnis Retail Hokimart, Kediri*. Institut Teknologi Sepuluh Nopember.
- Apriana, D., & Yuliansyah, C. (2024). Mengoptimalkan Penjualan *Online* Melalui Teknik Data Mining (Studi Kasus E-Commerce). *AL-MIKRAJ Jurnal Studi Islam Dan Humaniora (E-ISSN 2745-4584)*, 4(02), 514–527.
- 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.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 1165–1188.
- Dawis, A. M., Irmawati, H. A., Nur, N., Sulfoyanti, S. A. R., Wajidi, F., Muchtar, M., Firgiawan, W., Arifin, N., Insani, C. N., & Sinaga, F. M. (2022). *Data Mining Dan Manajemen Pengentahuan*.
- Dibb, S. (1998). Market segmentation: strategies for success. *Marketing Intelligence & Planning*, 16(7), 394–406.
- Gustiasari, N. S. (2024). *Strategi Komunikasi Pemasaran Digital PT Matahari Department Store Tbk dalam melakukan Personalisasi Pemasaran untuk menghadapi Persaingan Bisnis Retail*. Universitas Islam Indonesia.
- Haque, M. G., Iskandar, A. S., Erlangga, H., & Sunarsi, D. (2022). *Strategi pemasaran: Konsep, teori dan implementasi*.
- Hendriyani, Y. (2025). Data Mining BAB. *Dasar Dasar Data Mining: Konsep, Teknik Dan Aplikasi*, 130.
- Kumar, V., & Reinartz, W. (2016). Creating enduring customer value. *Journal of Marketing*, 80(6), 36–68.
- Pine, B. J., & Korn, K. C. (2011). *Infinite possibility: Creating customer value on the digital frontier*. Berrett-Koehler Publishers.
- Susilawati Sugiana, N. S., & Musty, B. (2023). Analisis Data Sistem Informasi Monitoring Marketing; Tools Pengambilan Keputusan Strategis. *Jutisi: Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*.
- Wafda, A. (2024). Integrasi Machine Learning dalam Ritel: Tinjauan Komprehensif tentang Prediksi Harga, Analisis Data Pelanggan, dan Pemanfaatan Media Sosial. *Journal Artificial: Informatika Dan Sistem Informasi*, 2(2), 90–106.
- Wicaksana, P. A., Swamardika, I. B. A., & Hartati, R. S. (2022). Literature review analisis perilaku pelanggan menggunakan rfm model. *Maj. Ilm. Teknol. Elektro*, 21(1), 21.
- Widyaningsih, M., & Rusdiana, L. (2025). Klasifikasi Citra Hewan Khas Suku Dayak Menggunakan Convolution Neural Network. *Sainteks*, 22(1), 87–98.
- Yusepa, W. Y., Kamilawati, A., Agesti, P. T., & Sanjaya, V. F. (2024). Peran Competitive Advantage Dalam Menarik Konsumen Di Pasar Lokal (Studi Pada Topshop Bandar Lampung). *Al-A'mal: Jurnal Manajemen Bisnis Syariah*, 1(2), 29–40.