Journal of Integrated System (JIS)
Vol. 8 No. 1 June 2025: 26-41
e-ISSN: 2621-7104

Received: 17 April 2025
Accepted: 19 June 2025
https://doi.org/10.28932/jis.v8i1.11503

Optimalisasi Strategi *Product Bundling* melalui Pemetaan Pola Peminatan dan Pola Penjualan Produk menggunakan *K-Means Clustering* dan *Apriori*

Optimizing Product Bundling Strategy through Mapping Product Interest Patterns and Sales Patterns using K-Means Clustering and Apriori

Bagas Dwi Oktavian, Ida Lumintu*, Samsul Amar
Program Studi Teknik Industri, Fakultas Teknik, Universitas Trunojoyo Madura, Bangkalan, Indonesia
*Penulis korespondensi, email: ida.lumintu@trunojoyo.ac.id

Abstrak

Persaingan bisnis ritel yang semakin kompetitif menuntut strategi pemasaran yang lebih terarah dan berbasis data untuk mengatasi stagnasi penjualan, khususnya pada produk dengan permintaan rendah. Penelitian ini bertujuan mengembangkan pendekatan analitik untuk mengoptimalkan strategi product bundling di toko Aldrin Stocklot's melalui pemetaan pola peminatan dan keterkaitan produk. Metode yang digunakan meliputi analisis Recency, Frequency, and Monetary (RFM) pada level produk, K-Means Clustering untuk segmentasi produk berdasarkan peminatan, dan algoritma Apriori untuk mengidentifikasi asosiasi antarproduk dalam transaksi pelanggan. Hasil menunjukkan bahwa 82% produk termasuk kategori kurang diminati, dan hanya 13 dari 33 aturan asosiasi yang efektif untuk bundling lintas klaster. Evaluasi clustering menggunakan Davies-Bouldin Index (DBI) sebesar 0.717 menandakan kualitas segmentasi yang cukup baik. Temuan ini memperkuat argumen bahwa integrasi klasterisasi dan association rule mining pada level produk dapat menghasilkan rekomendasi bundel yang lebih adaptif untuk meningkatkan visibilitas dan penjualan produk stagnan. Implikasi praktis dari pendekatan ini berpotensi direplikasi di sektor ritel UMKM lainnya.

Kata kunci: algoritma Apriori untuk penjualan, K-Means Clustering dalam bisnis, pola peminatan produk, pola penjualan ritel, strategi product bundling

Abstract

The increasing competitiveness in the retail business demands more targeted and data-driven marketing strategies to address stagnant sales, particularly for low-demand products. This study aims to develop an analytical approach to optimize product bundling strategies at Aldrin Stocklot's by mapping product preference patterns and inter-product associations. The methods used include Recency, Frequency, and Monetary (RFM) analysis at the product level, K-Means Clustering to segment products based on customer interest, and the Apriori algorithm to identify product associations within customer transactions. The results show that 82% of products fall into the low-interest category, and only 13 out of 33 association rules were found effective for cross-cluster bundling. Clustering performance evaluation using the Davies-Bouldin Index (DBI) yielded a score of 0.717, indicating a reasonably good segmentation quality. These findings support the argument that integrating clustering and association rule mining at the product level can generate more adaptive bundling recommendations to enhance the visibility and sales of stagnant products. The practical implications of this approach suggest it can be replicated in other small-scale retail businesses.

Keywords: Apriori algorithm for sales, K-Means Clustering in business, product bundling strategy, product preference pattern, retail sales pattern

1. Pendahuluan

Persaingan di sektor ritel semakin kompetitif seiring meningkatnya ekspektasi konsumen terhadap diferensiasi produk dan efisiensi layanan (Azemi and Ozuem, 2023; John and Thaiyalnayaki, 2024). Keberhasilan bisnis ritel dalam mempertahankan keberlanjutan penjualan sangat ditentukan oleh ketepatan strategi pemasaran yang digunakan untuk memahami dan mengakomodasi preferensi konsumen (Yin and Xu, 2021). Dalam konteks ini, salah satu tantangan utama dalam pengelolaan

How to Cite:
Oktavian, B.D., Lumintu, I. and Amar, S. (2025) 'Optimalisasi strategi product bundling melalui pemetaan pola peminatan dan pola penjualan produk menggunakan K-Means Clustering dan Apriori', *Journal of Integrated System*, 8(1), pp. 26–41. Available at: https://doi.org/10.28932/jis.v8i1.11503.

portofolio produk adalah disparitas minat pelanggan terhadap item yang tersedia. Produk dengan tingkat peminatan tinggi cenderung mendominasi transaksi, sedangkan produk lain stagnan dan memperlambat rotasi inventori (Ergin, Gümüş and Yang, 2021; Kim *et al.*, 2022; Rimbano *et al.*, 2023), sehingga menyebabkan inefisiensi operasional dan menurunkan potensi pendapatan (Rahmadsyah and Mayasari, 2022).

Strategi promosi konvensional seringkali belum mampu menjawab tantangan tersebut karena tidak dibangun di atas fondasi analisis pola konsumsi aktual. Dalam praktiknya, upaya untuk mendorong penjualan produk dengan minat rendah masih mengandalkan pendekatan generik yang tidak mempertimbangkan keterhubungan perilaku belanja konsumen (Martins *et al.*, 2021; Oetama, 2024). Hal ini menunjukkan perlunya pengembangan strategi yang didasarkan pada relasi fungsional antarproduk yang terekam dalam data transaksi historis (Chen, 2024; Hussein *et al.*, 2024; Ramadhan and Rokhman, 2020).

Penelitian terdahulu telah menunjukkan potensi besar dari integrasi antara teknik klasterisasi dan *association rule mining* untuk mengungkap pola laten dalam perilaku pembelian konsumen (Ardi, Nastiti and Sumarlinda, 2023; Sarkar, Puja and Chowdhury, 2024). Namun, mayoritas studi masih berfokus pada level pelanggan. Misalnya, Singh, Mittal and Pareek (2016), Husein *et al*. (2022), Yahya *et al.* (2024), dan Dey and Banerjee (2023) menerapkan pendekatan RFM dan algoritma *Apriori* untuk segmentasi pelanggan serta sistem rekomendasi berbasis segmen, tanpa mengkaji eksplisit keterkaitan antarproduk dalam konteks promosi *bundling*. Di sisi lain, Gustriansyah, Suhandi and Antony (2020) dan Agussalim, Wardhani and Anjani (2023) menghitung nilai RFM di level produk dan menerapkan klasterisasi, namun tidak melanjutkannya dengan pemodelan asosiasi maupun eksplorasi *bundling*. Bahkan ketika klasterisasi dan *Apriori* telah dikombinasikan seperti pada Chen and Gunawan (2023), Fernando *et al.* (2022), dan Rahmatillah, Sudirman and Sharif (2023), tujuan aplikasinya masih terbatas pada rekomendasi umum dan tidak menyasar pengembangan strategi *bundling* lintas klaster.

Penelitian ini mengusulkan sebuah *pipeline* berbasis data transaksi yang menyatukan analisis RFM produk, klasterisasi preferensi menggunakan pendekatan *unsupervised learning K-Means*, dan pemodelan asosiasi untuk menghasilkan bundel lintas klaster yang bertujuan meningkatkan visibilitas produk berdaya tarik rendah untuk mengisi *gap* dalam penelitian terdahulu. Tidak ditemukan *framework* sejenis dalam literatur sebelumnya yang menggabungkan ketiga komponen tersebut secara terintegrasi dalam konteks strategi *bundling* adaptif.

Urgensi dari pendekatan ini diperkuat oleh karakteristik kontekstual dari subjek studi. Aldrin Stocklot's, sebuah toko ritel berskala UMKM di Kabupaten Bangkalan, menjual lebih dari 180 variasi produk bayi dan balita. Berdasarkan data penjualan internal, lebih dari 60% item menunjukkan stagnasi permintaan dalam tiga bulan terakhir. Strategi promosi yang digunakan masih bersifat intuitif, tanpa mempertimbangkan disparitas performa produk atau keterhubungan antar item dalam transaksi. Kondisi ini mencerminkan tipikal masalah yang dihadapi banyak UMKM ritel dalam mengoptimalkan kombinasi produk dan pengalokasian upaya promosi secara efisien.

Secara akademik, kontribusi penelitian ini terletak pada perumusan *framework* metodologis yang menyatukan tiga komponen analitik pada tingkat produk—RFM-*based K-Means clustering*, *association rule mining*, dan seleksi bundel lintas klaster—yang belum terdokumentasi dalam penelitian sebelumnya (Izzah and Jananto, 2022; Rizaldi and Adnan, 2021; Susanto and Fenriana, 2024; Syrotkina, Bhatta and Jacob, 2023; Wahyuni, Wulansari and Fahrullah, 2023). Secara praktis, pendekatan ini membuka peluang formulasi strategi promosi yang lebih presisi dan berbasis data aktual, serta dapat direplikasi pada bisnis ritel berskala serupa. Dengan demikian, artikel ini tidak hanya memberikan kontribusi orisinal dalam tataran metodologis, tetapi juga menyajikan signifikansi praktis yang relevan untuk menjawab tantangan nyata dalam ekosistem ritel berbasis UMKM.

2. Metode

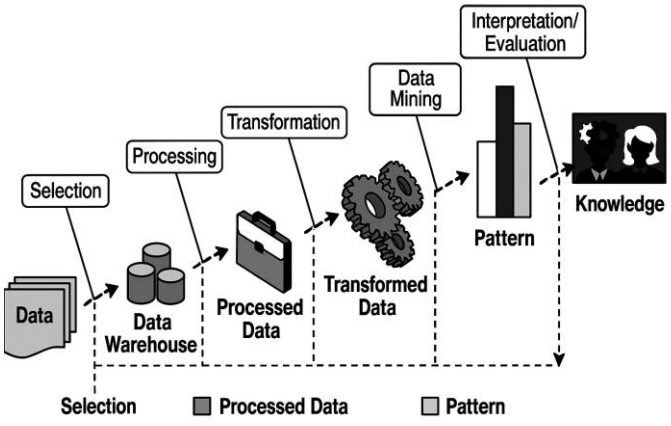
2.1 Tahap Pengambilan Data

Dalam penelitian ini, pengambilan data dilakukan melalui beberapa teknik yang dirancang untuk mendapatkan informasi yang relevan dengan tujuan penelitian. Teknik yang digunakan meliputi wawancara, observasi langsung, dan analisis data historis. Setiap metode memiliki peran penting dalam memberikan gambaran menyeluruh mengenai kondisi toko Aldrin Stocklot's, pola peminatan produk, serta keterkaitan antarproduk dalam transaksi penjualan.

1. Wawancara
   Wawancara dilakukan secara langsung dengan pemilik dan pegawai toko Aldrin Stocklot's untuk menggali informasi mengenai praktik manajemen toko serta strategi pemasaran yang telah diterapkan maupun yang belum pernah dicoba sebelumnya. Melalui wawancara ini, peneliti dapat memahami tantangan yang dihadapi dalam mengelola stok produk dan meningkatkan penjualan. Pendekatan wawancara yang digunakan bersifat terbuka, sehingga memungkinkan pemilik dan pegawai untuk memberikan wawasan yang lebih mendalam terkait pola penjualan dan preferensi pelanggan.
2. Observasi Langsung
   Selain wawancara, peneliti juga melakukan observasi langsung di toko Aldrin Stocklot's untuk mengumpulkan informasi mengenai pengelompokan dan penataan produk di toko. Observasi ini bertujuan untuk memahami bagaimana produk saat ini disusun dan bagaimana pelanggan berinteraksi dengan berbagai kategori produk yang tersedia. Dengan melihat kondisi langsung di lapangan, peneliti dapat mengidentifikasi faktor-faktor yang dapat memengaruhi efektivitas strategi *product bundling* yang akan diterapkan.
3. Analisis Data Historis
   Data historis yang digunakan dalam penelitian ini diperoleh dari pemilik toko Aldrin Stocklot's, mencakup transaksi penjualan selama periode satu tahun, yaitu dari 1 November 2023 hingga 31 Oktober 2024. Data ini mencakup waktu pembelian, kode transaksi, kode produk, nama produk, jumlah produk yang terjual, serta total pembayaran yang dilakukan oleh konsumen. Informasi dari data historis ini akan menjadi dasar utama dalam melakukan analisis pola peminatan pelanggan, menemukan keterkaitan antarproduk dalam transaksi, serta merancang strategi pemasaran yang lebih efektif melalui *product bundling*.

2.2 Tahap Penelitian

Penelitian ini menggunakan pendekatan *Knowledge Discovery in Database* (KDD) untuk menemukan pola dalam data transaksi toko Aldrin Stocklot's. KDD adalah suatu proses sistematis dalam mengeksplorasi data, yang terdiri dari lima tahapan utama: *Selection*, *Preprocessing*, *Transformation*, *Data Mining*, dan *Interpretation* (Ginantra *et al.*, 2021). Pendekatan ini memungkinkan analisis yang lebih terstruktur dalam mengekstrak informasi yang bernilai dari data penjualan yang tersedia.



Gambar 1. Proses KDD (Ginantra *et al.*, 2021)

2.2.1 Tahapan K-Means Clustering

Penelitian ini menggunakan metode *K-Means Clustering* untuk mengelompokkan produk berdasarkan pola peminatan pelanggan. Setiap tahap dalam metode ini mengikuti alur kerja KDD, yang dimulai dari pemilihan data hingga interpretasi hasil.

1. Data Selection

Data yang digunakan dalam penelitian ini merupakan data historis transaksi penjualan di toko Aldrin Stocklot's selama satu tahun, terhitung dari 1 November 2023 hingga 31 Oktober 2024. Data ini mencakup berbagai kategori produk, termasuk pakaian, aksesoris, dan minuman. Namun, dalam penelitian ini, hanya data pakaian dan aksesoris yang digunakan, sedangkan data produk minuman dikecualikan dari analisis.

Atribut-atribut dalam data transaksi mencakup berbagai informasi, seperti nomor transaksi, tanggal pembelian, kode pelanggan, nama pelanggan, alamat, kode item, nama item, jumlah pembelian, harga satuan, potongan harga, total pembayaran, pajak, biaya tambahan, dan total akhir. Dari berbagai atribut tersebut, hanya lima atribut utama yang digunakan dalam proses KDD untuk analisis pengelompokan produk:

a. Waktu Pemesanan: Mencatat waktu pembelian produk oleh pelanggan. Atribut ini digunakan untuk menganalisis pola pembelian berdasarkan waktu tertentu.
b. Kode Produk: Berisi kode unik yang merepresentasikan setiap produk di toko Aldrin Stocklot's.
c. Nama Produk: Menyimpan informasi mengenai nama produk yang tersedia di toko.
d. Jumlah Produk: Menunjukkan jumlah unit produk yang dibeli dalam satu transaksi.
e. Total Pembayaran: Mencatat jumlah pembayaran yang dilakukan pelanggan.

2. Preprocessing

Sebelum dilakukan proses *data mining*, data harus dibersihkan dari informasi yang tidak relevan, nilai yang hilang (*missing values*), dan data redundan. *Missing values* muncul ketika suatu entri kosong atau tidak memiliki nilai, sedangkan redundansi terjadi ketika data memiliki duplikasi yang tidak diperlukan. Proses *data cleaning* akan memastikan bahwa hanya data valid yang digunakan, sehingga analisis menjadi lebih akurat dan efisien. Menghilangkan *missing values* serta mengurangi redundansi meningkatkan keandalan model, memastikan hasil klasifikasi dan asosiasi produk dapat digunakan sebagai dasar strategi pemasaran yang lebih optimal.

3. Transformation

Tahap *transformation* mengubah data yang telah dipilih agar sesuai dengan proses *data mining*. Pada tahap ini, transformasi dilakukan menggunakan pendekatan *Recency, Frequency, and Monetary* (RFM). *Recency* dihitung berdasarkan selisih waktu antara tanggal transaksi terakhir suatu produk dengan periode penelitian, *frequency* diperoleh dari jumlah total pembelian produk dalam periode tertentu, dan *monetary* dihitung dari total nilai transaksi suatu produk (Prasetyo, Mustafid and Hakim, 2020). Setelah nilai RFM diperoleh, data kemudian dinormalisasi menggunakan *z-score normalization* untuk memastikan setiap variabel berada dalam skala yang seimbang sebelum proses *data mining* dilakukan.

$$z = \frac{x - mean}{stdev} \quad (1)$$

keterangan:
z : nilai hasil normalisasi
x : nilai asli yang diamati
mean : rata-rata dari seluruh nilai dalam dataset
stdev : standar deviasi dari dataset

Normalisasi *z-score* memastikan bahwa proses *K-Means Clustering* dapat berjalan lebih optimal tanpa terpengaruh oleh perbedaan skala antar variabel. Dengan normalisasi ini, jarak antar data menjadi lebih proporsional, menghindari dominasi variabel dengan nilai yang lebih besar. Setelah proses normalisasi

selesai, data yang telah disesuaikan digunakan sebagai *dataset* untuk tahap pemodelan menggunakan *K-Means Clustering* di *RapidMiner*.

4. K-Means Clustering

Pada tahap ini, *data mining* dilakukan dengan menggunakan metode *K-Means Clustering* yang merupakan proses utama dalam menemukan pola yang bermakna dari data yang telah dipilih dan diproses sebelumnya. Dalam penelitian ini, *K-Means Clustering* digunakan untuk mengelompokkan produk berdasarkan tingkat peminatannya. Tahapan dalam penerapan *K-Means Clustering* adalah sebagai berikut:

a. Menentukan jumlah cluster: Dalam penelitian ini, digunakan tiga cluster untuk mengelompokkan produk, yaitu *cluster 0*, *cluster 1*, dan *cluster 2*, yang masing-masing mewakili kategori *sangat diminati*, *diminati*, dan *kurang diminati*.
b. Memasukkan nilai normalisasi RFM: Data hasil normalisasi dari atribut *recency, frequency*, dan *monetary* dimasukkan ke dalam *RapidMiner* untuk proses analisis lebih lanjut.
c. Menyeleksi atribut yang digunakan: Hanya atribut yang relevan dipilih menggunakan operator *Select Attributes* di RapidMiner agar analisis lebih fokus dan efisien.
d. Menghitung jarak data dengan *centroid*: Perhitungan jarak antar data dilakukan menggunakan persamaan *Euclidean Distance* untuk menentukan kedekatan masing-masing data dengan pusat *cluster*:

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

keterangan:
x : titik data pertama
y : titik data kedua
n : jumlah karakteristik atau dimensi dalam data
$d(x,y)$: jarak *Euclidean* antara titik x dan y

e. *Clustering* dengan *K-Means*: Proses *clustering* dilakukan dengan $K = 3$ dan 10 iterasi untuk memastikan hasil yang optimal.
f. Penempatan data dalam cluster: Data ditempatkan ke dalam cluster terdekat berdasarkan *centroid* dengan optimasi menggunakan *performance vector*.
g. Evaluasi Davies-Bouldin *Index* (DBI): Hasil clustering dievaluasi menggunakan DBI, dimana nilai ≥ 0 menunjukkan clusterisasi yang baik. Perhitungan diawali dengan *Sum of Square Within-Cluster* (SSW) menggunakan nilai *centroid* pada iterasi terakhir (Kamalludin *et al.*, 2023):

$$SSW = \sum_{i=1}^{k}\sum_{j=1}^{m_i} d\left(x_j, c_i\right)^2 \quad (3)$$

keterangan:
m_i : jumlah data dalam *cluster* i
c_i : *centroid cluster ke-* i
$d(x_j, c_i)$: jarak antara data x_j dan *centroid* c_i
j : banyaknya data dalam *cluster*
x : data individu dalam *cluster*

SSW mengukur seberapa dekat data dalam satu *cluster* dengan *centroid*. Nilai SSW yang lebih kecil menunjukkan bahwa data dalam *cluster* lebih homogen dan hasil *clustering* lebih optimal. Setelah menghitung *Sum of Square Within-Cluster* (SSW), langkah berikutnya adalah menghitung *Sum of Square Between-Cluster* (SSB) untuk mengevaluasi pemisahan antar cluster. SSB mengukur sejauh mana setiap *centroid cluster* berbeda dari *centroid* keseluruhan, dengan nilai yang lebih besar menunjukkan pemisahan *cluster* yang lebih optimal. Perhitungan dilakukan menggunakan rumus berikut (Kamalludin *et al.*, 2023).

$$SSB = \sum_{i=1}^{k} m_i \cdot d(c_i, c_j)^2 \quad (4)$$

keterangan:
i, j : indeks *cluster*
c_i, c_j : *centroid* masing-masing *cluster*
$d(c_i, c_j)$: jarak antara *centroid cluster* i dan j
m_i : jumlah data dalam *cluster* i

Nilai SSB yang lebih besar menunjukkan pemisahan antar *cluster* yang lebih jelas, menandakan hasil *clustering* yang optimal. Setelah memperoleh nilai *Sum of Square Between-Cluster* (SSB), langkah berikutnya adalah menghitung rasio yang membandingkan pemisahan antara *cluster* i dan *cluster* j. Rasio ini digunakan untuk menilai kualitas *clustering* dengan melihat sejauh mana perbedaan antar *cluster* dibandingkan dengan kedalaman masing-masing *cluster*. Perhitungan pada tahap ini dilakukan menggunakan rumus berikut (Kamalludin *et al.*, 2023).

$$R_{(i,j)} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (5)$$

keterangan:
$R_{(i,j)}$: ukuran rasio yang mengevaluasi kualitas pemisahan antara *cluster* i dan *cluster* j
SSW_i : *Sum of Square Within-Cluster* untuk *cluster* i, yang mengukur variasi data dalam *cluster* tersebut.
SSW_j : *Sum of Square Within-Cluster* untuk *cluster* j, yang mengukur variasi data dalam *cluster* tersebut.
$SSB_{i,j}$: *Sum of Square Between-Cluster*, yang mengukur sejauh mana *cluster* i dan j terpisah berdasarkan jarak *centroid* mereka.

Nilai $R_{(i,j)}$ yang lebih kecil menunjukkan *cluster* yang lebih terpisah dengan baik, sedangkan nilai yang lebih besar menunjukkan bahwa *cluster* masih memiliki tumpang tindih atau tidak cukup terpisah. Langkah berikutnya adalah menghitung perbandingan antara *cluster* i dan *cluster* j untuk mengevaluasi kualitas hasil *clustering*. Evaluasi ini menggunakan Davies-Bouldin *Index* (DBI), dimana nilai DBI yang lebih rendah (non-negatif, ≥ 0) menunjukkan bahwa hasil *clustering* semakin baik. Semakin kecil nilai DBI, semakin jelas pemisahan antar *cluster* dan semakin homogen data dalam setiap *cluster*. Berikut adalah rumus akhir untuk menghitung Davies-Bouldin *Index* (Kamalludin *et al.*, 2023).

$$DBI = \frac{1}{k} \sum_{i=1}^{k} max_{i \neq j}(R_{i,j}) \quad (6)$$

keterangan:
DBI : Davies Bouldin *Index*
k : jumlah *cluster* yang digunakan
$R_{(i,j)}$: rasio antara kedalaman dalam *cluster* dan jarak antar *centroid cluster* i dan j

Nilai DBI yang lebih kecil menunjukkan *cluster* yang lebih terpisah dengan baik dan lebih homogen dalam masing-masing kelompok.

5. Interpretation / Evaluation

Hasil *data mining* disajikan dalam bentuk grafik dan plot *clustering* untuk memvisualisasikan pengelompokan produk sangat diminati, diminati, dan kurang diminati berdasarkan *K-Means Clustering*. Rata-rata nilai *Recency*, *Frequency*, dan *Monetary* (RFM) dalam setiap *cluster* turut dianalisis. Evaluasi *clustering* menggunakan *performance vector* dan Davies-Bouldin *Index* (DBI), dimana nilai DBI yang non-negatif dan lebih besar dari nol menunjukkan kualitas pengelompokan yang baik.

2.2.2 Tahapan Algoritma Apriori

Algoritma *Apriori* digunakan untuk menemukan pola penjualan berdasarkan transaksi pelanggan di toko Aldrin Stocklot's.

1. Data Selection

Data transaksi yang digunakan mencakup periode 1 November 2023 hingga 31 Oktober 2024, dengan fokus pada produk pakaian dan aksesoris, sementara data minuman dihapus dari analisis. Empat atribut utama yang digunakan adalah:

a. Jumlah Produk: Banyaknya unit produk yang dibeli.
b. Nama Produk: Identitas produk dalam transaksi.
c. Kode Produk: Kode unik untuk setiap produk.
d. Nomor Transaksi: Identifikasi satu transaksi pelanggan.

Data ini dipilih agar analisis dengan algoritma Apriori dapat mengungkap keterkaitan antar produk dalam pola pembelian pelanggan.

2. Preprocessing

Tahap *preprocessing* memastikan data bersih dan siap digunakan dengan menghilangkan data yang tidak relevan, *missing values*, dan redundansi. *Missing values* terjadi ketika suatu entri kosong atau tidak memiliki nilai, sementara data redundan adalah duplikasi yang tidak diperlukan. Pembersihan ini penting untuk menjaga akurasi dan validitas analisis.

3. Transformation

Tahap *transformation* bertujuan mengubah data yang telah dipilih agar sesuai dengan kebutuhan analisis. Proses ini dilakukan dengan menggunakan *pivot table* untuk mengidentifikasi produk yang dibeli dalam setiap nomor transaksi. Data kemudian disiapkan untuk dioperasikan di *RapidMiner* menggunakan algoritma Apriori. Dalam pemodelan *Apriori*, data dikonversi ke dalam format biner, dimana angka 1 menunjukkan bahwa produk dibeli, sementara angka 0 menunjukkan bahwa produk tidak dibeli. Hasil transformasi ini akan menjadi *dataset* yang digunakan sebagai input dalam proses analisis pola transaksi.

4. Algoritma Apriori

Tahap ini mencari pola bermakna dalam data berdasarkan alur *Knowledge Discovery in Database* (KDD). Proses dilakukan dengan algoritma *Apriori* untuk mengidentifikasi keterkaitan antar produk dalam transaksi pelanggan. Berikut tahapan dalam menentukan pola penjualan produk:

a. Menentukan minimum support dan iterasi kombinasi item set: Menetapkan nilai minimum support sebagai batas bawah frekuensi kemunculan item set, kemudian melakukan iterasi untuk mencari kombinasi item yang memenuhi ambang batas tersebut (Rizaldi and Adnan, 2021; Situmorang *et al.*, 2024).
b. Menentukan minimum support: Nilai minimum support ditentukan berdasarkan karakteristik data yang digunakan untuk memastikan hanya item dengan frekuensi kemunculan yang signifikan yang dipertimbangkan. Perhitungan support untuk satu item dilakukan menggunakan rumus:

$$\text{Support}\ (A) = \frac{\text{Jumlah transaksi mengandung } A}{\text{Total transaksi}} \quad (7)$$

Sedangkan untuk mencari nilai support dari dua item dalam satu transaksi digunakan rumus:

$$\text{Support}\ (A \cup B) = \frac{\sum \text{transaksi mengandung } A \text{ dan } B}{\text{Total transaksi}} \quad (8)$$

Dengan menentukan nilai minimum support yang tepat, hanya item set yang memiliki hubungan kuat dalam pola pembelian pelanggan yang akan dipertahankan untuk analisis lebih lanjut.
c. Menentukan minimum *confidence*: Nilai minimum *confidence* ditetapkan untuk mengidentifikasi aturan asosiasi yang memenuhi ambang batas kepercayaan tertentu. *Confidence* mengukur seberapa

besar kemungkinan item B muncul dalam transaksi ketika item A sudah ada (Rizaldi and Adnan, 2021; Situmorang *et al.*, 2024). Perhitungannya menggunakan rumus:

$$\text{Confidence} = P\,(B|A) = \frac{\sum \text{transaksi mengandung } A \text{ dan } B}{\sum \text{transaksi mengandung } A} \quad (8)$$

Dengan menetapkan minimum *confidence* yang sesuai, hanya aturan asosiasi yang memiliki hubungan kuat antar produk yang akan digunakan dalam analisis pola pembelian pelanggan (Situmorang *et al.*, 2024).

d. Memasangkan item terpilih dengan *RapidMiner*: Tahap akhir dilakukan dengan mengolah data menggunakan *RapidMiner* untuk menghasilkan kombinasi item yang memenuhi aturan asosiasi. Proses ini memungkinkan identifikasi pola penjualan berdasarkan keterkaitan antar produk, sehingga memberikan wawasan mengenai hubungan pembelian yang dapat digunakan untuk strategi pemasaran yang lebih efektif.

5. Interpretation/Evaluation

Pola penjualan yang dihasilkan dari *association rules* dalam algoritma *Apriori* dengan *RapidMiner*, yang didasarkan pada minimum *support* dan *confidence*, akan digunakan sebagai dasar strategi *product bundling* di toko Aldrin Stocklot's untuk meningkatkan efektivitas pemasaran.

3. Hasil dan Pembahasan

Hasil penelitian ini mencakup analisis peminatan pelanggan melalui pengelompokan produk dan identifikasi pola penjualan yang bertujuan mendukung strategi *product bundling*. Temuan-temuan tersebut memperkuat urgensi penerapan pendekatan promosi yang didasarkan pada pola konsumsi aktual, serta menunjukkan potensi strategis dari pendekatan bundling adaptif—suatu aspek yang hingga kini belum secara eksplisit dibahas dalam studi-studi sebelumnya.

3.1 K-Means Clustering

Setelah melalui tahap *preprocessing*, data ditransformasikan dengan menggunakan metode RFM untuk mempermudah proses pengolahan dalam *K-Means Clustering*. Tabel 1 menampilkan hasil transformasi data untuk setiap produk di toko Aldrin Stocklot's.

Tabel 1. Nilai *Recency*, *Frequency*, *Monetary* (RFM) produk

No.	Kode Produk	Nama Produk	*Recency*	*Frequency*	*Monetary*
1	754	SETELAN KIDS BOY	4	235	8480000
2	753	ST.KIDS BOY	2	435	14790000
3	715	SETELAN BABY BOY	4	140	7560000
4	560	GAMIS	360	2	124000
5	709	SET. BABY BOY	61	81	3564000
6	873	SET. BABY GIRL	7	80	5600000
7	48	KEMEJA LK	12	73	2920000
8	506	JAKET/SWEATER	163	53	2120000
9	6	KAOS BOY	26	49	1274000
10	1419	SOFA BABY	81	16	1280000
11	1300	FEEDING SET	19	54	1728000
12	342	CELANA PANJANG LK	16	117	5850000
13	1036	JILBAB	72	46	1012000
14	1041	JILBAB	82	105	3150000
15	858	STELAN PR KIDS	1	346	14532000
16	215	DRES BABY KIDS	14	235	10575000
17	864	SET KIDS GIRL	20	159	8268000
18	857	STELAN PR KIDS	1	119	4760000
19	806	SETELAN KIDS GIRL	20	51	1938000
...	...	...	...	...	...
822	1281	CELEMEK	1	1	15000

Langkah berikutnya adalah melakukan normalisasi nilai RFM menggunakan *z-score normalization* untuk menyeimbangkan skala data, memastikan *K-Means Clustering* lebih optimal, dan mempersiapkan *dataset* untuk pemodelan di *RapidMiner*. Tabel 2 menampilkan hasil normalisi nilai RFM. Setelah proses normalisasi nilai RFM, hasil *K-Means clustering* di *RapidMiner* menunjukkan bahwa dari total 822 produk di toko Aldrin Stocklot's, cluster 0 terdiri dari 675 produk, cluster 1 mencakup 7 produk, dan cluster 2 memiliki 140 produk seperti yang terlihat pada Gambar 2.

Visualisasi hasil clustering pada Gambar 3 menunjukkan pola peminatan produk berdasarkan analisis *K-Means Clustering* di *RapidMiner*. Dalam grafik hasil *clustering*, *cluster* 0 (biru) merepresentasikan produk yang kurang diminati, *cluster* 1 (hijau) menunjukkan produk yang sangat diminati, dan *cluster* 2 (merah) mencerminkan produk yang diminati. Grafik ini memberikan gambaran jelas mengenai hubungan antara *frequency* dan *monetary*, dimana produk dalam *cluster* 1 memiliki nilai tinggi, menunjukkan daya jual yang kuat, sementara produk dalam *cluster* 0 memiliki nilai paling rendah, mengindikasikan rendahnya tingkat pembelian. *Cluster* 2 berada di antara kedua kategori ini, menunjukkan bahwa produk dalam kelompok ini memiliki tingkat peminatan yang stabil, tetapi tidak sepopuler produk dalam *cluster* 1. Dengan memahami pola ini, strategi pemasaran yang lebih efektif, seperti *product bundling* atau promosi spesifik, dapat diterapkan untuk meningkatkan penjualan produk dalam *cluster* 0 dan 2. Tabel 3 menampilkan rincian hasil klasifikasi produk dalam *cluster* dari hasil *modeling* yang telah dilakukan.

Tabel 2. Normalisasi nilai RFM

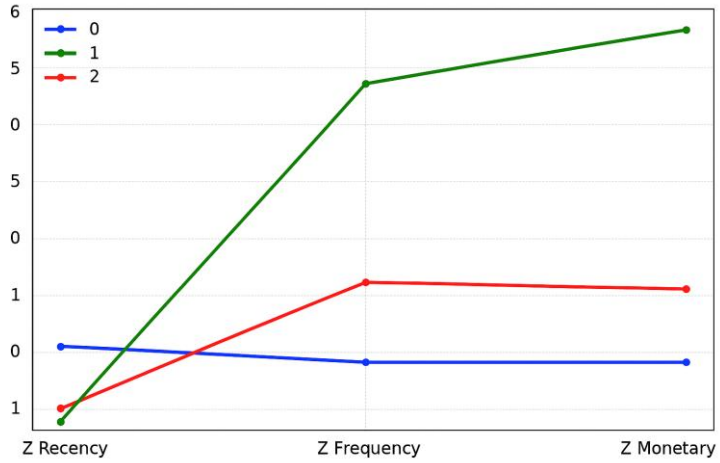
No.	Kode Produk	Nama Produk	*Z Recency*	*Z Frequency*	*Z Monetary*
1	754	SETELAN KIDS BOY	-0.9871	3.3443	2.7133
2	753	ST.KIDS BOY	-1.0077	6.7666	5.2072
3	715	SETELAN BABY BOY	-0.9871	1.7187	2.3497
4	560	GAMIS	2.6819	-0.6428	-0.5893
5	709	SET. BABY BOY	-0.3996	0.7091	0.7703
6	873	SET. BABY GIRL	-0.9561	0.6920	1.5750
7	48	KEMEJA LK	-0.9046	0.5722	0.5158
8	506	JAKET/SWEATER	0.6516	0.2299	0.1996
9	6	KAOS BOY	-0.7603	0.1615	-0.1348
10	1419	SOFA BABY	-0.1935	-0.4032	-0.1324
11	1300	FEEDING SET	-0.8325	0.2471	0.0447
12	342	CELANA PANJANG LK	-0.8634	1.3251	1.6738
13	1036	JILBAB	-0.2862	0.1102	-0.2383
14	1041	JILBAB	-0.1832	1.1198	0.6067
15	858	STELAN PR KIDS	-1.0180	5.2437	5.1053
16	215	DRES BABY KIDS	-0.8840	3.3443	3.5413
17	864	SET KIDS GIRL	-0.8222	2.0438	2.6295
18	857	STELAN PR KIDS	-1.0180	1.3593	1.2430
19	806	SETELAN KIDS GIRL	-0.8222	0.1957	0.1277
...	...	...	...	...	...
822	1281	CELEMEK	-1.0180	-0.6599	-0.6324

Cluster Model

```
Cluster 0: 675 items
Cluster 1: 7 items
Cluster 2: 140 items
Total number of items: 822
```

Gambar 2. Hasil *clustering* dengan *K-Means*

Pada Tabel 3, hasil *clustering* menunjukkan distribusi produk berdasarkan tingkat peminatannya di toko Aldrin Stocklot's. *Cluster* 0, yang terdiri dari 675 produk, mendominasi jumlah keseluruhan, mengindikasikan bahwa sebagian besar produk termasuk kategori kurang diminati. Sebaliknya, *cluster* 1 yang memiliki 7 produk, menunjukkan bahwa hanya sedikit produk yang sangat diminati pelanggan. *Cluster* 2 mencakup 140 produk, merepresentasikan kategori produk yang masih memiliki daya tarik, meskipun tidak sebesar *cluster* 1. Ketimpangan ini menunjukkan perlunya strategi pemasaran yang lebih agresif namun terarah, seperti misalnya *product bundling* atau promosi khusus, untuk meningkatkan daya jual produk dalam *cluster* 0 dan memperluas peminatan terhadap produk di *cluster* 2. Rata-rata nilai RFM dalam setiap cluster ditampilkan pada Tabel 4.



Gambar 3. Grafik *cluster*

Tabel 3. Rincian klasifikasi produk dalam *cluster*

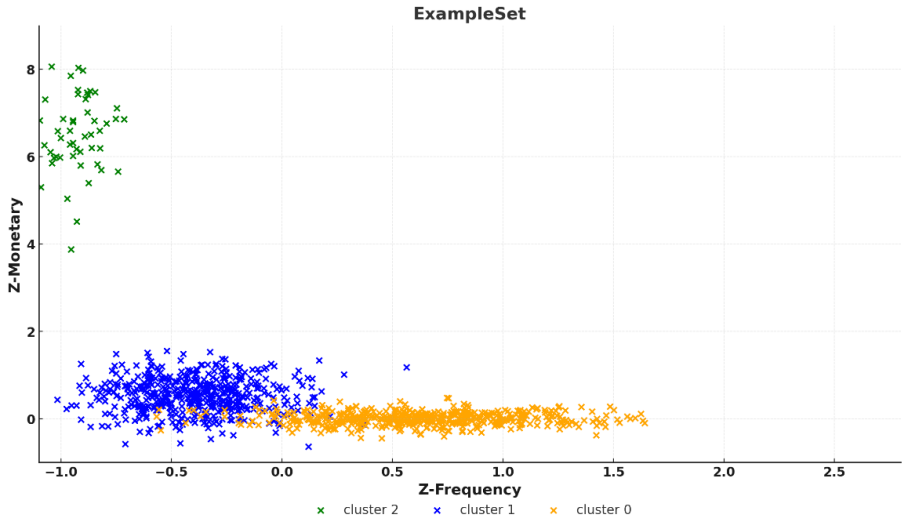
No.	Kode Produk	Nama Produk	Cluster
1	754	SETELAN KIDS BOY	2
2	753	ST.KIDS BOY	1
3	715	SETELAN BABY BOY	2
4	560	GAMIS	0
5	709	SET. BABY BOY	2
6	873	SET. BABY GIRL	2
7	48	KEMEJA LK	2
8	506	JAKET/SWEATER	0
9	6	KAOS BOY	0
10	1419	SOFA BABY	0
11	1300	FEEDING SET	0
12	342	CELANA PANJANG LK	2
13	1036	JILBAB	0
14	1041	JILBAB	2
15	858	STELAN PR KIDS	1
16	215	DRES BABY KIDS	2
17	864	SET KIDS GIRL	2
18	857	STELAN PR KIDS	2
19	806	SETELAN KIDS GIRL	0
...	...	...	...
822	1281	CELEMEK	0

Tabel 4. Rata-rata nilai RFM dalam setiap *cluster*

Cluster	R	F	M	Hasil
0	0.186	-0.366	-0.340	Kurang diminati
1	-0.710	5.339	6.264	Sangat diminati
2	-0.859	1.497	1.324	Diminati

Tabel 4 menunjukkan rata-rata nilai *Recency* (R), *Frequency* (F), dan *Monetary* (M) dalam setiap *cluster*, yang memberikan gambaran mengenai pola pembelian pelanggan. *Cluster* 0, yang dikategorikan sebagai kurang diminati, memiliki nilai *recency* tertinggi (0.186), menunjukkan bahwa produk dalam *cluster* ini relatif lebih baru dalam transaksi terakhir dibandingkan dengan *cluster* lainnya. Namun, nilai F (-0.366) dan M (-0.340) yang rendah menunjukkan bahwa meskipun produk ini masih muncul dalam transaksi, frekuensi pembelian dan total nilai penjualannya tetap rendah. Sebaliknya, *cluster* 1, yang termasuk dalam kategori sangat diminati, memiliki nilai F (5.339) dan M (6.264) yang jauh lebih tinggi dibandingkan dengan *cluster* lain, menunjukkan volume penjualan yang dominan. *Cluster* 2, yang mencerminkan produk diminati, memiliki nilai F (1.497) dan M (1.324) lebih tinggi dari *cluster* 0 tetapi lebih rendah dari *cluster* 1. Data ini menegaskan bahwa nilai *recency* yang tinggi tidak selalu mencerminkan performa penjualan yang baik, karena faktor utama dalam menentukan peminatan tetap bergantung pada frekuensi pembelian dan total nilai transaksi. Pada Gambar 4, grafik persebaran *cluster* memberikan visualisasi yang lebih jelas mengenai distribusi setiap *cluster* dari hasil *K-Means Clustering*.

Gambar 5 menunjukkan hasil *performance vector* dari proses *clustering*, yang mengevaluasi jarak rata-rata antara titik data dalam setiap cluster dan pusat (*centroid*). Nilai rata-rata jarak keseluruhan adalah 1.376, dengan *cluster* 0 memiliki jarak terendah (1.214), menunjukkan bahwa data dalam *cluster* ini lebih terpusat. Sebaliknya, *cluster* 1 memiliki jarak tertinggi (3.786), menandakan bahwa penyebaran data dalam kelompok ini lebih luas, yang dapat mengindikasikan adanya variasi tinggi dalam peminatan produk. *Cluster* 2 memiliki nilai 2.038, yang berada di antara kedua *cluster* lainnya. Indeks Davies-Bouldin (DBI) sebesar 0.717 menunjukkan bahwa hasil *clustering* cukup baik, karena nilai yang lebih rendah mengindikasikan pemisahan antar *cluster* yang lebih jelas. Dengan DBI yang non-negatif ≥ 0, *clustering* yang diperoleh dapat dikatakan cukup optimal, meskipun masih ada potensi penyempurnaan untuk meningkatkan pemisahan antar kelompok produk.



Gambar 4. Distribusi *cluster*

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 1.376
Avg. within centroid distance_cluster_0: 1.214
Avg. within centroid distance_cluster_1: 3.786
Avg. within centroid distance_cluster_2: 2.038
Davies Bouldin: 0.717
```

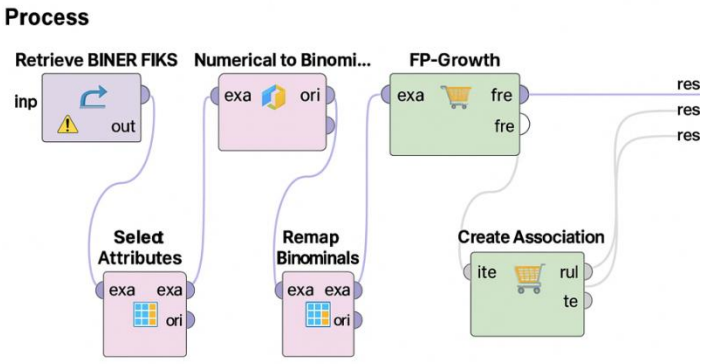
Gambar 5. Nilai *performance vector*

3.2 Algoritma Apriori

Data yang akan diproses dalam algoritma *Apriori* dipilih berdasarkan lima atribut utama yaitu nomor transaksi, nama produk, kode produk, serta jumlah produk yang dibeli. Setelah data diseleksi, dilakukan *preprocessing* untuk memastikan data bersih sebelum masuk ke tahap transformasi. Transformasi ini dilakukan menggunakan *pivot table* di *Excel* untuk mengidentifikasi kode transaksi dalam setiap pembelian. Dalam pemodelan Apriori, data kemudian dikonversi ke format biner (0 dan 1), dimana nilai 1 menunjukkan produk dibeli, dan nilai 0 menunjukkan produk tidak dibeli. Hasil transformasi ini disiapkan untuk digunakan dalam *RapidMiner* dan dapat dilihat pada Tabel 5. Data kemudian dinormalisasi menggunakan *RapidMiner* sebelum diproses dalam algoritma *Apriori* untuk memastikan hasil analisis yang optimal (Gambar 6).

Tabel 5. Hasil transformasi dalam format biner

KODE TRANSAKSI	STELAN PR KIDS=856	STELAN PR KIDS=857	STELAN PR KIDS=858	STELAN PR KIDS=860	STELAN PR KIDS=867	STELAN PR KIDS=868	.....	TALI RAMBUT=1361
1	0	0	0	0	0	0		0
2	0	0	0	0	0	0		0
3	0	0	0	0	0	0		0
4	0	0	0	0	0	0		0
7	0	0	0	0	0	0		0
8	0	0	0	0	0	0		0
9	0	0	0	0	0	0		0
10	0	0	0	0	0	0		0
12	0	0	0	0	0	0		0
13	0	1	1	0	0	0		0
14	0	0	0	0	0	0		0
15	0	0	0	0	0	0		0
16	0	0	0	0	0	0		0
17	0	0	0	0	0	0		0
18	0	0	0	0	0	0		0
19	0	1	0	0	0	0		0
21	0	0	0	0	0	0		0
22	0	0	0	0	0	0		0
23	0	0	0	0	0	0		0
24	0	0	0	0	0	0		0
25	0	0	1	0	0	0		0
26	0	0	0	0	0	0		0
27	0	0	0	0	0	0		0
28	0	0	0	0	0	0		0
29	0	0	0	0	0	0		0
30	0	0	0	0	0	0		0
31	0	0	0	0	0	0		0
32	0	0	0	0	0	0		0
33	0	0	0	0	0	0		0
34	0	0	0	0	0	0		0
35	0	0	0	0	0	0		0
36	0	0	0	0	0	0		0
38	0	0	0	0	0	0		0
.....								
16794	0	0	0	0	0	0		0



Gambar 6. Algoritma *Apriori*

Proses ini dimulai dengan mengimpor *file Excel*, menyeleksi atribut menggunakan *Select Attributes*, dan mengonversi data numerik ke format binominal melalui *Numerical to Binominal*. Selanjutnya, nilai 0 (tidak dibeli) dan 1 (dibeli) disesuaikan menggunakan *Remap Binominal* agar data dapat diproses dengan benar dalam model asosiasi. Pada Gambar 6, untuk menemukan pola keterkaitan antar produk dalam penelitian ini, proses *FP-Growth* menggunakan *minimum support* sebesar 0.03% dan proses *Create Association Rules* menggunakan *minimum confidence* sebesar 5%. Pemilihan nilai *minimum support* dalam penelitian ini mempertimbangkan besarnya jumlah transaksi, sehingga pola yang muncul dalam persentase kecil tetap diperhitungkan tanpa mengabaikan keterkaitan yang berpotensi signifikan. Sementara itu, nilai *minimum confidence* 5% ditetapkan agar aturan asosiasi yang terbentuk memiliki tingkat kepercayaan yang cukup tinggi tanpa terlalu membatasi jumlah pola yang ditemukan. Kombinasi parameter ini memastikan keseimbangan antara relevansi aturan yang diperoleh dan jumlah pola yang dapat digunakan dalam analisis lebih lanjut.

Hasil *association rules* dari algoritma *Apriori* menghasilkan 33 kombinasi produk terbaik, yang mencakup kombinasi dari produk dalam *cluster* yang sama maupun berbeda. Karena itu, tidak semua kombinasi ternyata efektif dalam meningkatkan penjualan produk dengan tingkat peminatan lebih rendah, karena beberapa kombinasi hanya terdiri dari produk dalam *cluster* yang sama sehingga kurang mendukung strategi peningkatan penjualan. Oleh karena itu, penelitian ini merekomendasikan 13 kombinasi produk yang lebih optimal seperti yang terlihat di Tabel 5, dengan fokus untuk mendorong produk dalam *cluster* peminatan lebih rendah agar dapat meningkatkan volume penjualannya. Rekomendasi ini dirancang sebagai strategi *product bundling* yang bisa diterapkan di toko Aldrin Stocklot's untuk memperkuat efektivitas pemasaran dan mendorong peningkatan penjualan.

Tabel 5. Rekomendasi *product bundling*

No.	*Premises*	*Cluster*	*Conclusion*	*Cluster*	*Support*	*Confidence*
1	DRES BABY KIDS=215	2	DRESS BABY-KIDS=220	1	0.09%	6.13%
2	DRES BABY KIDS=223	2	DRES BABY KIDS=219	1	0.06%	6.47%
3	DRESS BABY-KIDS=214	2	DRESS BABY-KIDS=220	1	0.04%	6.74%
4	DRESS BABY-KIDS=217	2	DRESS BABY-KIDS=220	1	0.08%	6.79%
5	DRES BABY KIDS=222	2	DRESS BABY-KIDS=220	1	0.06%	7.26%
6	CELANA PJ GIRL=416	0	ATASAN GIRL=143	2	0.03%	11.36%
7	BANDO=1422	0	BANDO=1024	2	0.04%	15.00%
8	CELANA DALAM=915	0	CELANA DALAM=912	2	0.03%	16.67%
9	CINCIN=1505	0	CINCIN=1111	2	0.08%	17.19%
10	DRES BABY KIDS=209	0	DRES BABY KIDS=211	2	0.06%	18.60%
11	CELANA DALAM=911	0	CELANA DALAM=913	2	0.03%	19.23%
12	CELANA PD GIRL=1350	0	ATASAN GIRL=142	2	0.05%	26.92%
13	CELANA PD GIRL=1350	0	ATASAN GIRL=141	2	0.06%	30.77%

4. Simpulan

Penelitian ini mengusulkan pendekatan berbasis data untuk meningkatkan volume penjualan di toko Aldrin Stocklot's dengan menganalisis pola peminatan produk dan keterkaitan produk dalam transaksi pelanggan. Dengan menerapkan *K-Means Clustering*, produk diklasifikasikan ke dalam tiga kategori utama: sangat diminati, diminati, dan kurang diminati. Hasil analisis menunjukkan bahwa mayoritas produk (675 dari 822 produk) masuk dalam kategori kurang diminati, yang mengindikasikan perlunya strategi pemasaran yang lebih efektif untuk meningkatkan daya tariknya. Sementara itu, hanya 7 produk berada dalam kategori sangat diminati, sedangkan 140 produk masuk dalam kategori diminati, mencerminkan adanya celah yang perlu dioptimalkan dalam strategi penjualan.

Lebih lanjut, penerapan algoritma *Apriori* berhasil mengidentifikasi 33 kombinasi produk berdasarkan pola keterkaitan dalam transaksi pelanggan. Namun, tidak semua kombinasi terbukti efektif dalam meningkatkan penjualan produk dengan tingkat peminatan rendah, karena beberapa kombinasi hanya menghubungkan produk dalam klaster yang sama tanpa memberikan dampak signifikan terhadap

perputaran stok. Oleh karena itu, penelitian ini merekomendasikan 13 kombinasi produk yang lebih optimal, yang dirancang sebagai strategi *product bundling* untuk meningkatkan eksposur dan daya tarik produk dengan permintaan rendah melalui kombinasi dengan produk yang lebih populer.

Evaluasi performa clustering menggunakan *Davies-Bouldin Index* (DBI) sebesar 0.717 menunjukkan bahwa pemisahan antar klaster cukup baik, meskipun masih terdapat ruang untuk perbaikan dalam meningkatkan kejelasan segmentasi produk. Sementara itu, pemilihan nilai *minimum support* dan *confidence* dalam algoritma *Apriori* memastikan bahwa aturan asosiasi yang dihasilkan memiliki relevansi yang kuat dalam mendukung strategi pemasaran toko.

Penelitian ini memiliki beberapa keterbatasan**.** Pertama, data yang digunakan terbatas pada satu toko ritel dengan skala UMKM dan periode waktu satu tahun, sehingga hasilnya belum dapat digeneralisasi ke jenis ritel lain atau konteks yang lebih luas. Kedua, analisis tidak mempertimbangkan faktor eksternal seperti promosi musiman, tren pasar, atau perilaku pesaing, yang dapat memengaruhi pola pembelian. Ketiga, metode *clustering* yang digunakan masih berbasis partisi statis dan belum dievaluasi terhadap model alternatif seperti DBSCAN atau *hierarchical clustering*.

Secara praktis, penelitian ini memberikan panduan konkret bagi pelaku usaha ritel untuk mengidentifikasi produk dengan performa rendah dan merancang strategi *bundling* berbasis data transaksi yang aktual. Pendekatan ini dapat digunakan untuk meningkatkan efisiensi pengelolaan stok, mengurangi penumpukan barang tidak laku, dan merancang promosi yang lebih terarah sesuai dengan preferensi pelanggan yang tersirat dalam data pembelian.

Dari sisi pengembangan ilmu pengetahuan, penelitian ini memperluas pemanfaatan algoritma *K-Means* dan *Apriori* dengan menggabungkannya dalam satu *framework* yang berfokus pada level produk, bukan pelanggan, untuk menyusun strategi *bundling* lintas klaster. Integrasi metodologis ini membuka ruang bagi penelitian lanjutan yang dapat menggabungkan faktor eksternal, pendekatan *dynamic clustering*, atau sistem rekomendasi yang lebih kompleks untuk menyempurnakan pendekatan pemasaran berbasis data dalam ekosistem UMKM ritel.

Daftar Pustaka

Agussalim, Wardhani, R.K.P. and Anjani, A. (2023). Sales product clustering using RFM calculation model and K-Means algorithm on Primskystore. *Technium Romanian Journal of Applied Sciences and Technology*, 16, pp.176–182. doi:https://doi.org/10.47577/technium.v16i.9978.

Ardi, R.B., Nastiti, F.E. and Sumarlinda, S. (2023). Algoritma K-Means Clustering untuk segmentasi pelanggan (studi kasus: Fashion Viral Solo). *Infotech journal*, 9(1), pp.124–131. doi:https://doi.org/10.31949/infotech.v9i1.5214.

Azemi, Y. and Ozuem, W. (2023). How does retargeting work for different Gen Z mobile users? *Journal of Advertising Research*, 63(4), pp.2023-023. doi:https://doi.org/10.2501/jar-2023-023.

Chen, A.H.-L. and Gunawan, S.T.D. (2023). Enhancing retail transactions: a data-driven recommendation using modified RFM analysis and association rules mining. *Applied sciences*, 13(18), pp.10057–10057. doi:https://doi.org/10.3390/app131810057.

Chen, Q. (2024). Application of K-Means algorithm in marketing. *Advances in Economics Management and Political Sciences*, 71(1), pp.178–184. doi:https://doi.org/10.54254/2754-1169/71/20241485.

Dey, D. and Banerjee, K. (2023). AI driven customer segmentation and recommendation of product for super mall. *Journal of Mines, Metals and Fuels*, 71(5), pp.656–660. doi:https://doi.org/10.18311/jmmf/2023/34166.

Ergin, E., Gümüş, M. and Yang, N. (2021). An empirical analysis of intra-firm product substitutability in fashion retailing. *Production and Operations Management*, 31(2), pp.607–621. doi:https://doi.org/10.1111/poms.13569.

Fernando, A.M.P., Adhikari, A.M.T.T., Wijesekara, W.H.A.T.K., Vithanage, T.V.T.I., Gamage, A. and Jayalath, T. (2022). Planning marketing strategies in small-scale business using data analysis. *2022 3rd International Informatics and Software Engineering Conference (IISEC), Informatics and*

Software Engineering Conference (IISEC), 2022 3rd International, pp.1–6. doi:https://doi.org/10.1109/IISEC56263.2022.9998200.

Ginantra, N. L. W. S. R. *et al.* (2021) *Data mining dan penerapan algoritma.* 1st edn.. Yayasan Kita Menulis.

Gustriansyah, R., Suhandi, N. and Antony, F. (2020). Clustering optimization in RFM analysis based on K-Means. *Indonesian Journal of Electrical Engineering and Computer Science*, 18(1), p.470. doi:https://doi.org/10.11591/ijeecs.v18.i1.pp470-477.

Husein, A.M., Setiawan, D., Sumangunsong, A.R.K., Simatupang, A. and Yasmin, S.A. (2022). Combination grouping techniques and association rules for marketing analysis based customer segmentation. *SinkrOn*, 7(3), pp.1998–2007. doi:https://doi.org/10.33395/sinkron.v7i3.11571.

Hussein, R.M., Khaw, K.W., Gaber, A. and Chew, X. (2024). Examining the behaviors and preferences of online shopping customers using clustering techniques. *Journal of Soft Computing and Data Mining*, 5(1). doi:https://doi.org/10.30880/jscdm.2024.05.01.009.

Izzah, L. and Jananto, A. (2022). Penerapan algoritma K-Means Clustering untuk perencanaan kebutuhan obat di Klinik Citra Medika. *Progresif: Jurnal Ilmiah Komputer*, 18(1), p.69. doi:https://doi.org/10.35889/progresif.v18i1.769.

John, S. and Thaiyalnayaki, M. (2024). Customer satisfaction in retail services – a study with reference to Kottayam District in Kerala. *Salud Ciencia y Tecnología - Serie de Conferencias*, 3, pp.907–907. doi:https://doi.org/10.56294/sctconf2024907.

Kamalludin, A., Budianto, U., Sakti, D.V.S.Y. and Christian, J. (2023). Penerapan algoritma klasterisasi K-Means kinerja produk dengan analisis Recency, Frequency, Monetary pada Kafe XYZ. *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, [online] 2(1), pp.258–266. Available at: https://senafti.budiluhur.ac.id/senafti/article/view/631 [Accessed 20 Mar. 2025].

Kim, J., Lee, C., Lee, W.-H., Ok, Y. and Thi, T. (2022). Idiosyncratic volatility, turnover and the cross-section of stock returns: evidence from the Korean stock market. *International Journal of Emerging Markets*, 18(12), pp.6192–6213. doi:https://doi.org/10.1108/ijoem-09-2021-1499.

Martins, P., Rodrigues, P., Martins, C., Barros, T., Duarte, N., Dong, R.K., Liao, Y., Comite, U. and Yue, X. (2021). Preference between individual products and bundles: Effects of complementary, price, and discount level in Portugal. *Journal of Risk and Financial Management*, 14(5), p.192. doi:https://doi.org/10.3390/jrfm14050192.

Oetama, R. (2024). Product bundling strategy for office supplies retailer through association rules mining: Comparative study of Apriori and ECLAT algorithms. *Indonesian Journal of Computer Science*, 12(6). doi:https://doi.org/10.33022/ijcs.v12i6.3516.

Prasetyo, S.S., Mustafid, M. and Hakim, A.R. (2020). Penerapan fuzzy C-Means kluster untuk segmentasi pelanggan e-commerce dengan metode Recency, Frequency, Monetary (RFM). *Jurnal Gaussian*, 9(4), pp.421–433. doi:https://doi.org/10.14710/j.gauss.v9i4.29445.

Rahmadsyah, A. and Mayasari, N. (2022). Grouping goods with the association rule method using A priori algorithms in modern retail stores. *International Journal Of Science Technology & Management*, 3(6), pp.1527–1532. doi:https://doi.org/10.46729/ijstm.v3i6.637.

Rahmatillah, I., Sudirman, I.D. and Sharif, O.O. (2023). Unveiling consumer behavior patterns using K-Medoids and association rule: a study on a medium-sized grocery store. *2023 International Conference on Digital Business and Technology Management (ICONDBTM)*. doi:https://doi.org/10.1109/icondbtm59210.2023.10327325.

Ramadhan, D.R. and Rokhman, N. (2020). Segmentation-based sequential rules for product promotion recommendations as sales strategy (case study: Dayra Store). *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 14(3), p.243. doi:https://doi.org/10.22146/ijccs.58107.

Rimbano, D., Marlita, D., Yuniarty, M., Anggraini, D. and Andrinaldo, A. (2023). Effectiveness and efficiency of working capital use in industrial sector companies exchange in *12th Year Proceedings International Conference on Business Economics & Management*, pp.566–578. doi:https://doi.org/10.47747/icbem.v1i1.1275.

Rizaldi, D. and Adnan, A. (2021). Market basket analysis menggunakan algoritma Apriori: kasus transaksi 212 Mart Soebrantas Pekanbaru. *Jurnal Statistika dan Aplikasinya*, 5(1), pp.31–40. doi:https://doi.org/10.21009/jsa.05103.

Sarkar, M., Puja, A.R. and Chowdhury, F.R. (2024). Optimizing marketing strategies with RFM method and K-Means Clustering-based AI customer segmentation analysis. *Journal of Business and Management Studies*, [online] 6(2), pp.54–60. doi:https://doi.org/10.32996/jbms.2024.6.2.5.

Singh, J., Mittal, M. and Pareek, S. (2016). Customer behavior prediction using K-Means Clustering Algorithm. in: M. Mittal and N.H. Shah, eds., *Optimal Inventory Control and Management Techniques*. IGI Global Scientific Publishing, pp.256–267. doi:https://doi.org/10.4018/978-1-4666-9888-8.ch013.

Situmorang, B.H., Isra, A., Paragya, D. and Adhieputra, D.A.A. (2024). Apriori algorithm application for consumer purchase patterns analysis. *Komputasi: Jurnal Ilmiah Ilmu Komputer dan Matematika*, 21(1), pp.15–20. doi:https://doi.org/10.33751/komputasi.v21i1.9260.

Susanto, J. and Fenriana, I. (2024). Web-based POS with Apriori Market Basket Analysis. *bit-Tech*, [online] 6(3), pp.271–280. doi:https://doi.org/10.32877/bt.v6i3.926.

Syrotkina, O., Bhatta, S. and Jacob, K. (2023). Recency-Frequency-Monetary analysis and recommendation system using Apriori algorithm on e-commerce sales data. *2023 13th International Conference on Advanced Computer Information Technologies (ACIT)*. doi:https://doi.org/10.1109/acit58437.2023.10275513.

Wahyuni, S., Wulansari, T.T. and Fahrullah (2023). Segmentasi pelanggan berdasarkan analisis Recency, Frequency, Monetary menggunakan algoritma K-Means pada CV. Toedjoe Sinar Group. *JURTI (Jurnal Rekayasa Teknologi Informasi)*, 7(2), pp.180–180. doi:https://doi.org/10.30872/jurti.v7i2.8748.

Yahya, M., Parenreng, J.M., Fathahillah, F., Wahid, A., Wahid, M.S.N. and Fajar B, M. (2024). A machine learning model for local market prediction using RFM model. *Elinvo (Electronics, Informatics, and Vocational Education)*, 9(1), pp.24–37. doi:https://doi.org/10.21831/elinvo.v9i1.58671.

Yin, W. and Xu, B. (2021). Effect of online shopping experience on customer loyalty in apparel business-to-consumer ecommerce. *Textile Research Journal*, 91(23-24), p.004051752110165. doi:https://doi.org/10.1177/00405175211016559.