

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Artikel Penelitian

Integrasi Algoritma FP-Growth dan K-Means untuk Analisis Keranjang Belanja dan Segmentasi Pelanggan pada Data Transaksi Ritel

^aShafa Salsabila, ^aIka Novita Dewi*

^aProgram Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang 50131

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 14 Juli 2025

Revisi Akhir: 30 Agustus 2025

Diterbitkan Online: 01 September 2025

KATA KUNCI

Transaksi Retail,
FP-Growth,
K-Means,
Segmentasi Pelanggan,
Keranjang Belanja

KORESPONDENSI

E-mail: ikadewi@dsn.dinus.ac.id*

A B S T R A C T

Seiring meningkatnya volume transaksi dalam industri ritel, kebutuhan untuk memahami perilaku konsumen secara mendalam menjadi semakin krusial. Penelitian ini bertujuan untuk menerapkan analisis keranjang belanja dan segmentasi pelanggan guna mengidentifikasi pola pembelian produk sekaligus memahami karakteristik pelanggan berdasarkan perilaku transaksional. Dataset yang digunakan adalah *Retail Transaction Dataset* dari *Kaggle*, berisi satu juta transaksi ritel tahun 2020 hingga 2024. Dataset diproses melalui tahapan pembersihan, transformasi format data, dan seleksi transaksi yang memenuhi kriteria minimal pembelian 2 produk per transaksi. Analisis dilakukan dengan algoritma FP-Growth dan K-Means. Hasil analisis FP-Growth menunjukkan adanya 16 aturan asosiasi dengan nilai support tertinggi sebesar 0.31%, confidence 7.35%, dan lift 0.89, dengan produk *Toothpaste* menjadi produk yang paling sering diasosiasikan. Segmentasi pelanggan dilakukan menggunakan algoritma K-Means dengan atribut demografis pelanggan, yaitu *Payment Method*, *Customer Category*, *Promotion*, *Season*, *Discount Applied*, dan *Store Type*. Jumlah kluster optimal ditentukan menggunakan Metode Elbow dan Davies-Bouldin Index (DBI). Hasil menunjukkan $k = 3$ sebagai jumlah kluster terbaik dengan nilai DBI sebesar 2.4724, yang menandakan pemisahan kluster cukup baik. Berdasarkan hasil segmentasi K-Means, diperoleh tiga kluster pelanggan dengan karakteristik berbeda, yaitu *retiree*, *teenager*, dan *profesional*. Pengelompokan kluster ini mencerminkan kecenderungan kategori dominan yang muncul pada atribut *Customer Category*. Integrasi kedua metode ini menghasilkan rekomendasi strategi pemasaran berbasis segmen yang lebih personal, seperti *bundling* sederhana untuk pelanggan *retiree*, promosi visual dan hadiah menarik untuk pelanggan *teenager*, dan sistem poin loyalitas untuk pelanggan *profesional*. Temuan ini diharapkan membantu pelaku industri ritel memahami perilaku pelanggan secara lebih mendalam dan menyusun strategi pemasaran yang lebih tepat sasaran.

1. PENDAHULUAN

Industri ritel di Indonesia mengalami pertumbuhan yang meningkat dalam beberapa tahun terakhir. Hal ini tercermin dalam data Badan Pusat Statistik yang hingga tahun 2021, jumlah gerai ritel modern berbasis minimarket, supermarket, dan hypermarket telah melampaui dari 40.000 unit [1]. Hal ini menunjukkan adanya pergeseran preferensi konsumen yang mengutamakan kenyamanan dan efisiensi dalam berbelanja. Pergeseran ini membuat transaksi harian di industri ritel modern semakin padat, ditandai dengan peningkatan omzet tahunan yang

mencapai 10-15% [2]. Namun, di balik tingginya volume transaksi, banyak pelaku usaha ritel yang belum memanfaatkan data tersebut secara optimal untuk meningkatkan strategi pemasaran dan rekomendasi produk [3], [4].

Data transaksi harian seperti pola belanja pelanggan, kecenderungan pembelian produk tertentu, hingga respons pelanggan terhadap promosi sering kali tidak dianalisis secara efektif [5]. Akibatnya, strategi pemasaran cenderung bersifat umum dan tidak berbasis perilaku konsumen yang aktual [6]. Penutupan permanen gerai Giant pada tahun 2021 oleh PT Hero Supermarket merupakan contoh nyata dari kegagalan strategi

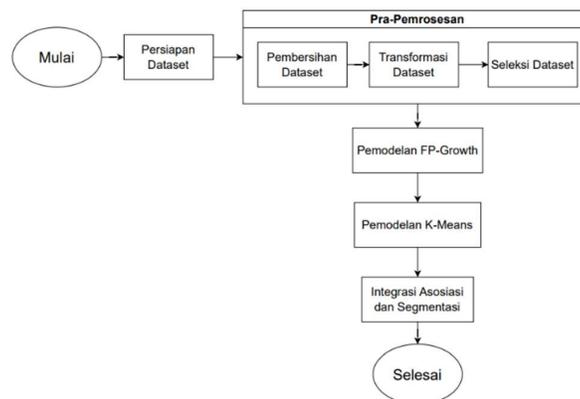
pemasaran yang adaptif terhadap dinamika pola pembelian dan permintaan pasar [7]. Kasus tersebut menunjukkan bahwa tanpa dukungan analitik yang kuat, perusahaan ritel rentan terhadap perubahan dinamika pasar.

Pemanfaatan teknik analitik data merupakan strategi yang kuat bagi industri ritel dalam merumuskan keputusan dan strategi bisnis [8]. Salah satu teknik analitik yang dapat digunakan adalah analisis aturan asosiasi. Analisis aturan asosiasi adalah teknik yang digunakan untuk mengidentifikasi hubungan antar produk dalam suatu set transaksi [9]. Salah satu metode yang dapat diterapkan dalam analisis asosiasi adalah algoritma FP-Growth. Algoritma FP-Growth digunakan untuk mengidentifikasi keterkaitan antar produk berdasarkan pola belanja serta memahami karakteristik umum yang muncul dari produk tersebut [10] [11] [12]. Sehingga dengan adanya penerapan teknik analitik data ini diharapkan industri ritel mendapatkan informasi untuk membangun sistem rekomendasi produk yang bersifat personal. Agar hasil rekomendasi sesuai dengan karakteristik pelanggan yang beragam, analisis segmentasi pelanggan juga perlu dilakukan menggunakan metode K-Means [13]. Dengan algoritma K-Means pelanggan dapat dikelompokkan berdasarkan kesamaan pola belanja sehingga industri ritel dapat merancang strategi pemasaran yang sesuai dengan preferensi tiap segmen [14].

Penelitian tentang analisis data berbasis transaksi pelanggan telah dilakukan pada beberapa penelitian sebelumnya. Seperti, Zahrudin (2024) [15] menganalisis tren penjualan fashion import dengan algoritma FP-Growth yang berhasil menemukan 33 aturan asosiasi dengan tingkat confidence 100% yang memberikan wawasan untuk merancang strategi penjualan di Toko Air Gaul. Dio (2023) [16] mengimplementasikan algoritma FP-Growth untuk analisis market basket pada transaksi belanja di retail, yang menghasilkan 4 aturan asosiasi dengan lift ratio > 1 , serta memberikan rekomendasi tata letak toko untuk meningkatkan efisiensi pembeli saat berbelanja.

Penelitian oleh Febrianty (2023) [17] mengaplikasikan algoritma K-Means untuk mengelompokkan pelanggan di sektor ritel online, menghasilkan empat segmen berdasarkan pola pembelian, harga unit, dan ID pelanggan untuk perancangan strategi pemasaran yang lebih efisien. Putri (2024) [18] mengkaji penggunaan algoritma K-Means untuk menganalisis data penjualan di Toko Ritel *Accessories*, yang menghasilkan lima kelompok produk berdasarkan tingkat minat beli sehingga dapat membantu toko dalam merancang strategi pemasaran dan mengelola stok produk lebih efisien. Sementara itu, Pertiwi (2024) [19] mengaplikasikan algoritma FP-Growth dan K-Medoids untuk menemukan asosiasi produk serta segmentasi pelanggan menjadi lima segmen mencerminkan tingkat loyalitas dan pola pembelian. Namun, penelitian-penelitian ini cenderung fokus pada asosiasi produk atau segmentasi pelanggan secara terpisah, sehingga belum menggabungkan kedua aspek tersebut secara menyeluruh. Meskipun terdapat penelitian yang telah mengintegrasikan keduanya, karakteristik pelanggan dan pola pembelian produk tidak terfokus pada kebutuhan spesifik tiap segmen. Sehingga kurangnya personalisasi dalam rekomendasi produk dan strategi pemasaran.

Penelitian ini mengusulkan model analisis yang mengintegrasikan algoritma FP-Growth dan K-Means untuk menghasilkan sistem rekomendasi produk yang lebih personal serta strategi pemasaran yang lebih tepat sasaran. Model ini ditetapkan melalui beberapa langkah, yaitu persiapan dan pra-pemrosesan data transaksi, penerapan FP-Growth untuk menemukan pola asosiasi antar produk, penerapan K-Means untuk segmentasi pelanggan berdasarkan pola pembelian, serta integrasi hasil keduanya untuk menghasilkan rekomendasi produk sesuai dengan karakteristik tiap segmen pelanggan. Pengujian model ini didasarkan pada kebutuhan industri ritel yang tidak hanya mampu memahami pola pembelian produk, tetapi juga melihat perbedaan karakteristik dan preferensi pelanggan. Melalui model ini, penelitian diharapkan dapat membantu industri ritel memahami pelanggan secara lebih mendalam sehingga dapat mendorong pertumbuhan penjualan.



Gambar 1. Diagram Alur Penelitian

2. METODE

Penelitian ini dilakukan dengan tujuan untuk mengembangkan model analisis yang dapat mengidentifikasi pola pembelian

<https://doi.org/10.25077/TEKNOSI.v11i2.2025.128-135>

produk sekaligus memahami karakteristik pelanggan di industri ritel. Untuk mencapai tujuan tersebut, beberapa tahapan penelitian perlu dilakukan, meliputi persiapan dataset, pra-pemrosesan, pemodelan FP-Growth, pemodelan K-Means,

serta integrasi asosiasi dan segmentasi. Tahapan penelitian ini dapat dilihat pada Gambar 1.

2.1. Dataset

Tahap awal dalam penelitian ini adalah persiapan dataset, yang menjadi dasar penting sebelum melakukan analisis data. Dataset yang digunakan adalah Retail Transaction Dataset [20], data transaksi penjualan ritel yang memuat informasi perilaku pembelian pelanggan.

Dataset yang digunakan memiliki 1.000.000 catatan transaksi penjualan ritel selama periode tahun 2020 hingga 2024 dengan 10 atribut, antara lain: *ID Transaction, Date, Product, Total cost, Payment Method, Customer Category, Promotion, Season, Discount Applied, Store Type*. Dengan variasi data yang cukup luas, dataset ini menjadi landasan yang kuat untuk dilakukan analisis pola pembelian serta segmentasi pelanggan. Sebagai ilustrasi, dataset ditampilkan dalam Tabel 1.

Tabel 1. Contoh Dataset

ID Transaction	Date	Product	...	Store Type
0	2022-01-21	['Ketchup', 'Shaving Cream', 'Light Bulbs']	...	Warehouse Club
1	2023-03-01	['Ice Cream', 'Milk', 'Olive Oil', 'Bread', 'P...']	...	Specialty Store
2	2024-03-21	['Spinach']	...	Department Store
3	2020-10-31	['Tissues', 'Mustard']	...	Pharmacy
...
999999	2020-06-15	['Trash Cans', 'Mop', 'Jam']	...	Pharmacy

2.2. Pra-Pemrosesan Data

Pra-pemrosesan data merupakan tahapan krusial untuk memastikan kualitas dan kesiapan data sebelum dianalisis lebih lanjut [21]. Dalam penelitian ini, pra-pemrosesan data dilakukan dengan pembersihan data, transformasi data, dan seleksi data.

2.2.1. Pembersihan Data

Tahapan ini difokuskan untuk menangani data yang memiliki nilai kosong pada seluruh atribut penting. Tujuannya untuk menjaga kualitas data dan memastikan semua data tetap dapat digunakan dalam analisis. Pada atribut *promotion* ditemukan nilai kosong sebanyak 181.389 baris. Mengingat proporsinya yang signifikan serta relevansinya terhadap analisis promosi, nilai kosong pada atribut ini tidak diabaikan. Dalam penelitian ini, nilai kosong tersebut diisi dengan label “*No Promotion*” untuk mewakili transaksi yang tidak memiliki penawaran promosi apapun. Langkah ini bertujuan agar data tetap utuh tanpa kehilangan informasi penting yang berkaitan dengan pola promosi. Sebagai ilustrasi, Tabel 3 menampilkan cuplikan data sebelum dan sesudah proses pembersihan pada atribut *promotion*.

Tabel 3. Data Sebelum dan Sesudah Pembersihan

ID	Promotion (Sebelum)	Promotion (Sesudah)
0	NaN	No Promotion
1	BIG1 (Buy One Get One)	BIG1 (Buy One Get One)
2	NaN	No Promotion
...
999999	Discount on Selected Items	Discount on Selected Items

2.2.2. Transformasi Data

Setelah proses pembersihan, langkah selanjutnya adalah melakukan transformasi data agar formatnya sesuai dengan kebutuhan analisis, khususnya untuk algoritma FP-Growth. Transformasi difokuskan pada kolom *product*, yang sebelumnya masih tersimpan dalam bentuk string list diubah menjadi list Python yang sesungguhnya. Hasil transformasi data dapat dilihat pada Tabel 4 berikut.

Tabel 4. Hasil Transformasi Kolom Produk dari Format String ke List

Sebelum Transformasi	Sesudah Transformasi
['Ketchup', 'Shaving Cream', 'Light Bulbs']	['Ketchup', 'Shaving Cream', 'Light Bulbs']
['Ice Cream', 'Milk', 'Olive Oil', 'Bread', 'P...']	['Ice Cream', 'Milk', 'Olive Oil', 'Bread', 'P...']
['Spinach']	['Spinach']

Selain itu, dilakukan transformasi data untuk mendukung segmentasi pelanggan menggunakan algoritma K-Means. Atribut-atribut pada data pelanggan seperti *Payment Method, Customer Category, Promotion, Season, Discount Applied*, dan *Store Type* yang bersifat kategorikal dikonversi menjadi format numerik menggunakan teknik One-Hot Encoding. Proses ini penting agar data dapat dibaca dan diproses oleh algoritma K-Means yang hanya menerima input numerik.

Melalui teknik One-Hot Encoding, setiap nilai kategori dari atribut tersebut dipecah menjadi kolom-kolom baru yang merepresentasikan keberadaan kategori tersebut dalam bentuk nilai biner, yaitu 0 dan 1. Misalnya, atribut *Payment Method* yang memiliki tiga kategori seperti *Cash, Credit Card*, dan *Mobile Payment* akan menghasilkan tiga kolom baru bernama *Payment_Method_Cash, Payment_Method_Credit Card*, dan *Payment_Method_Mobile Payment*. Nilai 1 menandakan bahwa transaksi tersebut menggunakan metode pembayaran yang dimaksud, sedangkan 0 berarti tidak. Bentuk data ini menjadi representasi numerik dari karakteristik pelanggan yang siap digunakan dalam proses klusterisasi, dan hasil transformasi ini dapat dilihat pada Tabel 5.

Tabel 5. Hasil Transformasi teknik One-Hot Encoding

ID Transaction	Payment_Method_Cash	...	Store_Type_Warehouse Club
0	0	...	1
1	1	...	0
2	0	...	0
...
999998	0	...	0

2.2.3. Seleksi Data

Tahap ini bertujuan untuk menyeleksi transaksi yang memenuhi syarat minimum untuk dianalisis menggunakan algoritma asosiasi, yaitu transaksi yang melibatkan setidaknya dua produk dalam satu pembelian. Hal ini penting karena algoritma FP-Growth tidak dapat membentuk pola asosiasi jika transaksi hanya berisi satu item, sebab tidak ada kombinasi produk yang bisa dianalisis.

Secara teknis, seleksi dilakukan dengan menyaring transaksi yang memuat dua atau lebih produk dalam satu pembelian. Dari total 1.000.000 transaksi, sebanyak 434.391 transaksi memenuhi kriteria ini dan dipertahankan untuk proses analisis lebih lanjut.

2.3. Pemodelan FP-Growth

Tahapan ini dilakukan untuk menemukan asosiasi produk, yaitu pola hubungan antar produk yang sering dibeli secara bersamaan oleh pelanggan. Untuk melakukan analisis ini, digunakan algoritma FP-Growth, yaitu salah satu metode dalam data mining yang digunakan untuk menemukan kombinasi item yang sering muncul dalam suatu kumpulan transaksi [22]. FP-Growth dipilih karena kemampuannya yang cepat dan efisien dalam menangani data transaksi yang besar [23].

FP-Growth menghasilkan aturan-aturan asosiasi produk dalam bentuk *if-then*, seperti "jika pelanggan membeli produk A, maka kemungkinan besar juga membeli produk B". Untuk mengevaluasi kualitas setiap aturan, digunakan tiga metrik utama:

1. Support menggambarkan seberapa sering kombinasi produk A dan B muncul bersama dalam seluruh transaksi. Semakin tinggi nilai support, semakin umum pola tersebut ditemukan. Support dapat dihitung dengan persamaan (1).

$$\text{Support}(A, B) = \frac{\text{Jumlah transaksi A dan B}}{\text{Jumlah total transaksi}} \times 100\% \quad (1)$$

2. Confidence menunjukkan seberapa besar kemungkinan konsumen membeli produk B jika sudah membeli produk A. Nilai confidence membantu memahami kekuatan hubungan sebab-akibat antar produk. Confidence dapat dihitung dengan persamaan (2).

$$\text{Confidence}(A, B) = \frac{\text{Jumlah transaksi A dan B}}{\text{Jumlah transaksi A}} \times 100\% \quad (2)$$

3. Lift Ratio mengukur hubungan antara produk A dan B lebih kuat dibandingkan dengan hubungan acak. Lift > 1 menunjukkan bahwa pembelian produk A dapat meningkatkan kemungkinan pembelian produk B. Lift Ratio dapat dihitung dengan persamaan (3).

$$\text{Lift}(A, B) = \frac{\text{Support}(A, B)}{\text{Support A} \times \text{Support B}} \quad (3)$$

2.4. Pemodelan K-Means

Tahap ini merupakan proses penerapan algoritma K-Means untuk membentuk model pengelompokan data berdasarkan kemiripan karakteristik. K-Means adalah algoritma klusterisasi yang bekerja dengan membagi data ke dalam sejumlah kluster berdasarkan

jarak terdekat antara masing-masing data dengan dengan pusat kluster (centroid) [24]. Algoritma ini dipilih karena memiliki keunggulan dalam efisiensi komputasi, khususnya saat digunakan pada data berdimensi tinggi dengan jumlah kluster yang relatif sedikit [25].

Proses pemodelan K-Means dilakukan melalui tahapan berikut:

1. Menentukan jumlah kluster optimal dengan menggunakan metode Elbow. Pada metode ini, dilakukan perbandingan Sum of Square Error (SSE) berbagai nilai k mulai dari nilai 2 hingga 9, dan nilai inerti dihitung untuk masing-masing nilai k. Metode ini mencari jumlah kluster optimal dengan mengamati titik di mana penurunan nilai SSE mulai melambat secara signifikan. Titik ini ditandai dengan "siku" pada grafik, yaitu saat penurunan SSE tidak lagi tajam meskipun jumlah kluster bertambah. Rumus SSE dihitung menggunakan persamaan (4).

$$SSE = \sum_{i=1}^n \sum_{x \in I} |x_i - c_k|^2 \quad (4)$$

2. Mengevaluasi pemilihan jumlah cluster menggunakan Davies-Bouldin Index (DBI). DBI mengukur seberapa baik pemisahan cluster yang terbentuk. Semakin rendah nilai DBI, semakin baik kualitas kluster yang dihasilkan. DBI dihitung berdasarkan persamaan (5).

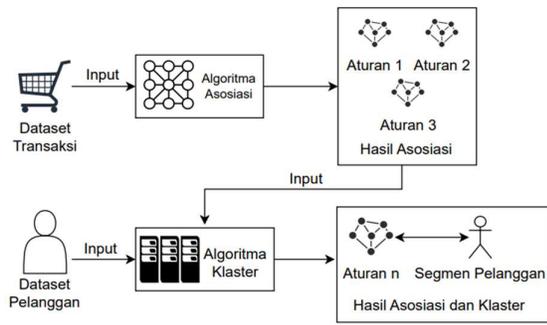
$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{i,j}) \quad (5)$$

3. Setelah nilai k ditentukan, maka menghasilkan sejumlah segmen pelanggan yang memiliki pola dan karakteristik belanja yang serupa. Setiap segmen kemudian dianalisis lebih lanjut untuk memahami preferensi, perilaku, dan profil khas masing-masing segmen pelanggan yang terbentuk.

2.5. Integrasi Hasil Aturan Asosiasi dan Segmentasi

Tahapan ini akan menggabungkan hasil dari dua proses analisis sebelumnya, yaitu analisis asosiasi produk menggunakan algoritma FP-Growth dan segmentasi pelanggan melalui pemodelan kluster menggunakan algoritma K-Means. Tujuannya adalah untuk memperoleh pemahaman mengenai hubungan antara pola pembelian produk dan karakteristik kelompok pelanggan tertentu. Dengan demikian, rekomendasi produk dan strategi pemasaran yang diberikan dapat disesuaikan dengan kebutuhan dan kebiasaan belanja masing-masing segmen.

Tahapan integrasi dilakukan dengan mencocokkan data transaksi yang memenuhi aturan asosiasi, yakni kombinasi produk yang sering muncul bersamaan dengan kluster pelanggan yang terbentuk melalui algoritma K-Means. Melalui tahap ini, setiap aturan asosiasi dapat dikaitkan dengan segmen pelanggan dominan yang sering terlibat dalam pola pembelian tersebut. Maka analisis tidak hanya dilakukan pada tingkat produk, tetapi dalam konteks perilaku pelanggan sehingga dapat menjadi landasan kuat untuk rekomendasi produk dan strategi pemasaran. Ilustrasi tahapan integrasi ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan integrasi hasil aturan asosiasi ke dalam proses segmentasi pelanggan

3. HASIL

3.1. Hasil Pemodelan FP-Growth

Beberapa kombinasi parameter minimum support dan minimum confidence diuji untuk mendapatkan aturan asosiasi yang optimal. Hasil percobaan menunjukkan variasi jumlah aturan yang dihasilkan sesuai dengan perubahan parameter. Tabel 5 menyajikan hasil perbandingan jumlah aturan asosiasi berdasarkan kombinasi parameter minimum support dan minimum confidence yang telah diuji.

Tabel 5. Perbandingan Jumlah Aturan Asosiasi

Min Support	Min Confidence	Jumlah Aturan	Lift Tertinggi
0.003	0.07	5	0.89
0.003	0.05	5	0.89
0.002	0.05	80	0.89
0.002	0.07	16	0.89

Berdasarkan hasil pada tabel 5, pemilihan kombinasi parameter minimum support 0.002 dan minimum confidence 0.07 dengan jumlah 16 aturan dipilih untuk tahap selanjutnya karena menghasilkan jumlah aturan yang cukup banyak namun tetap relevan. Adapun aturan asosiasi dapat dilihat pada Tabel 6.

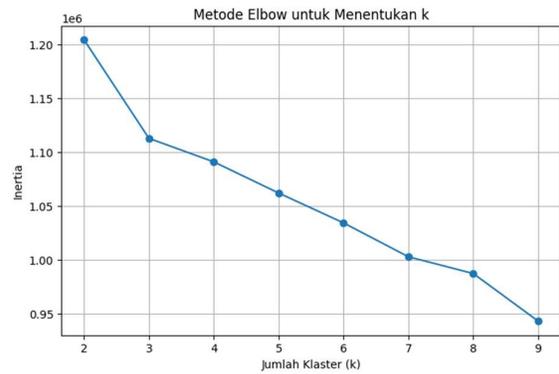
Tabel 6. Aturan Asosiasi Produk

Antecedents	Consequent	Support	Confidence	Lift
Apple	Toothpaste	0.31	7.35	0.89
Baby Wipes	Toothpaste	0.31	7.29	0.88
Cheese	Toothpaste	0.30	7.21	0.87
Olive Oil	Toothpaste	0.30	7.24	0.87
Toilet Paper	Toothpaste	0.30	7.16	0.87
Trash Cans	Toothpaste	0.30	7.16	0.87
Yogurt	Toothpaste	0.30	7.11	0.86
Ketchup	Toothpaste	0.30	7.09	0.86
Deodorant	Toothpaste	0.30	7.12	0.86
Cleaning	Toothpaste	0.30	7.15	0.86
Spray				
Broom	Toothpaste	0.29	7.05	0.85
Onions	Toothpaste	0.30	7.01	0.85
Air Freshener	Toothpaste	0.30	7.04	0.85
Cleaning Rags	Toothpaste	0.30	7.07	0.85
Hand Sanitizer	Toothpaste	0.29	7.00	0.85
Rice	Toothpaste	0.29	7.03	0.85

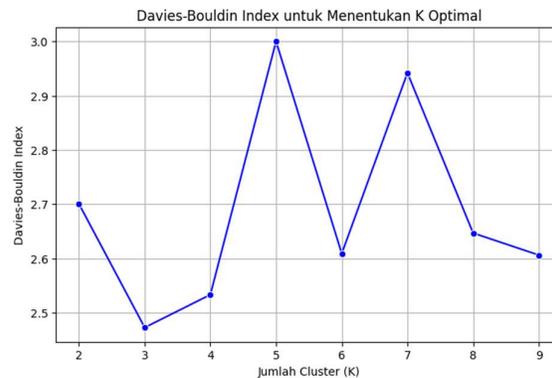
Aturan-aturan pada tabel 6 menunjukkan kombinasi produk yang paling sering dibeli bersamaan oleh pelanggan dalam dataset transaksi yang dianalisis.

3.2. Hasil Pemodelan K-Means

Penentuan jumlah cluster optimal dilakukan menggunakan dua metode evaluasi, yaitu Metode Elbow dan DBI. Grafik Elbow pada Gambar 3 menunjukkan bahwa titik tekuk atau elbow point berada pada nilai $k = 3$, yang menandakan bahwa menambahkan jumlah kluster lebih dari tiga tidak memberikan penurunan inersia yang signifikan. Hal ini didukung oleh perhitungan DBI pada Gambar 4, di mana titik DBI paling rendah tercatat pada $k = 3$, mengindikasikan pemisahan kluster yang cukup baik.



Gambar 2. Grafik Metode Elbow



Gambar 3. Grafik Metode Davies-Bouldin Index (DBI)

Setelah jumlah kluster ditentukan, hasil klusterisasi dengan K-Means menghasilkan tiga kluster pelanggan dengan proporsi yang relatif seimbang. Hasil analisis dapat dilihat pada Tabel 7.

Tabel 7. Hasil Analisis Kluster

Cluster	Member	Persentase (%)	Rata-Rata Total Transaksi
0	333.625	33.36	52.48
1	332.687	33.27	52.42
2	333.688	33.37	52.47

Setelah menentukan jumlah kluster optimal sebanyak tiga kluster, dilakukan analisis lebih mendalam untuk menggambarkan karakteristik khas dari masing-masing kluster yang terbentuk. Analisis ini bertujuan untuk memahami profil dominan dari tiap

segmen pelanggan berdasarkan atribut-atribut pada data transaksi. Ringkasan karakteristik setiap segmen ditampilkan pada Tabel 8. Setiap atribut pada tabel merepresentasikan nilai modus atau nilai yang paling sering muncul dalam klaster tersebut.

Tabel 8. Karakteristik Klaster

Cluster	Payment Method	Customer Category	Store Type
0	Credit Card	Retiree	...	Pharmacy
1	Cash	Teenager	...	Supermarket
2	Mobile Payment	Professional	...	Supermarket

Berdasarkan tabel 8, dapat dijabarkan karakteristik masing-masing klaster sebagai berikut:

1. Cluster 0 didominasi oleh pelanggan berasal dari kategori *Retiree* dengan metode pembayaran *Credit Card*, dan cenderung berbelanja saat musim *Spring* dengan memanfaatkan *Discount on Selected Items*. Transaksi pada klaster ini sebagian besar terjadi di *Pharmacy* dan seringkali mendapatkan diskon.
2. Cluster 1 terdiri dari pelanggan berusia lebih muda, yaitu *Teenager* dengan metode pembayaran *Cash*. Klaster ini sering merespons promosi *BOGO (Buy One Get One)*, serta didominasi pembelian pada musim *Spring* di Supermarket. Klaster ini sering memanfaatkan diskon.
3. Cluster 2 terdiri dari profil pelanggan *Professional* yang cenderung menggunakan *Mobile Payment*. Promosi yang sering dimanfaatkan adalah *Discount on Selected Items*, namun pembelian pada klaster ini banyak terjadi di musim *Fall* dan sebagian besar di Supermarket. Sebagian besar transaksi di klaster ini tidak menggunakan diskon.

3.3. Hasil Integrasi Asosiasi Produk dan Segmentasi Pelanggan

Sebagai hasil integrasi, Tabel 9 menampilkan tiga aturan asosiasi produk dengan nilai support, confidence, dan lift tertinggi yang mewakili masing-masing segmen pelanggan yang terbentuk. Setiap asosiasi yang ditampilkan merupakan kombinasi produk yang paling kuat di dalam satu segmen pelanggan dominan.

Tabel 9. Aturan Asosiasi Produk Tiap Segmen

Antecedents	Consequent	Support	Confidence	Lift
Apple	Toothpaste	0.003062	0.0735	0.8884
Cleaning Spray	Toothpaste	0.002988	0.0715	0.8644
Cheese	Toothpaste	0.003041	0.0721	0.8715

Dari hasil dalam tabel 9, produk *Toothpaste* secara konsisten muncul sebagai produk *consequent* dalam setiap asosiasi, yang berarti *Toothpaste* memiliki kecenderungan kuat untuk dibeli bersamaan dengan produk lain seperti *Apple*, *Cleaning Spray*, dan *Cheese*. Nilai support yang diperoleh cukup stabil, mencerminkan bahwa kombinasi produk ini sering terjadi dalam data transaksi. Meskipun nilai confidence yang diperoleh berada di kisaran 7% dan lift masih di bawah angka 1, aturan-aturan ini tetap signifikan karena dianalisis dari dataset yang besar,

sehingga meskipun persentase kecil, mencakup ribuan transaksi. Dalam hal ini dapat terlihat pada tabel 10.

Tabel 10. Segmentasi Pelanggan Dominan

Dominan Cluster	Cluster Dominance (%)	Total Matched Transaction
0	33.71	35.748
1	33.47	36.185
2	33.65	35.975

Hasil integrasi yang ditampilkan dalam Tabel 10 menunjukkan bahwa setiap asosiasi produk melibatkan jumlah transaksi yang cukup besar, yaitu sekitar 35.000 transaksi. Klaster 0 memiliki proporsi terbesar yaitu 33.71%, diikuti oleh Klaster 2 sebesar 33.65%, dan Klaster 1 sebesar 33,47%.

4. PEMBAHASAN

4.1. Analisis Temuan Penelitian

Berdasarkan hasil pengolahan dataset transaksi industri ritel menggunakan analisis asosiasi dan segmentasi pelanggan, penelitian ini menyajikan sejumlah temuan yang dapat dijadikan dasar untuk merumuskan rekomendasi produk dan strategi pemasaran di sektor ritel. Strategi yang diusulkan fokus pada pendekatan promosi yang relevan dengan pola pembelian pelanggan, seperti bundling produk, sistem poin loyalitas, serta penawaran diskon yang disesuaikan dengan kebiasaan masing-masing segmen. Beberapa strategi dapat dirancang secara spesifik untuk masing-masing klaster, sebagaimana dijelaskan berikut ini:

1. Pelanggan pada klaster 0 berasal dari kategori *Retiree* yang cenderung melakukan pembelian secara terencana dan mengutamakan kenyamanan. Metode pembayaran yang digunakan adalah *Credit Card*, dengan lokasi belanja utama di *Pharmacy*, serta kecenderungan berbelanja saat musim *Spring*. Promosi yang disukai berupa *Discount on Selected Items*, menunjukkan sensitivitas terhadap harga. Aturan asosiasi yang dominan adalah *Apple* → *Toothpaste*, yang mengindikasikan pola pembelian untuk kebutuhan dapur dan kebersihan. Berdasarkan karakteristik ini, strategi promosi yang paling sesuai adalah bundling praktis, yaitu paket dua produk yang relevan secara fungsi, ditawarkan dengan potongan harga tetap yang mudah dipahami, seperti "Beli *Apple* dan *Toothpaste*, hemat Rp2.000.". Strategi ini dipilih karena pelanggan lansia umumnya menghindari skema promosi yang rumit atau berbasis poin, dan lebih menyukai penawaran langsung yang sederhana serta mudah diakses. Penempatan bundling ini sebaiknya dilakukan di area yang strategis seperti rak dekat kasir atau lorong farmasi agar mudah dijangkau oleh *Retiree*.
2. Pelanggan pada klaster 1 didominasi oleh *Teenager* yang melakukan pembayaran dengan *Cash*, sering berbelanja di *Supermarket*, dan sangat responsif terhadap promosi *BOGO (Buy One Get One)*. Pelanggan remaja biasanya menunjukkan perilaku impulsif, mudah tergoda promosi visual, dan tertarik pada penawaran tambahan. Hasil asosiasi menunjukkan pola *Cleaning Spray* → *Toothpaste*, yang dapat

dijadikan dasar untuk promosi kombinasi produk rumah tangga ringan. Strategi yang disarankan adalah bundling dua produk berbeda dengan hadiah langsung atau promosi kreatif lainnya, seperti “Beli *Cleaning Spray* + *Toothpaste*, gratis stiker karakter”. Penempatan promosi di lokasi strategis dan visual yang menarik akan memperkuat efek promosi bagi segmen ini.

- Pelanggan pada klaster 2 merupakan profesional yang menunjukkan kecenderungan belanja efisien dan rasional, dengan preferensi metode *Mobile Payment*, sedikit ketergantungan pada diskon, dan kebiasaan belanja di *Supermarket* pada musim *Fall*. Asosiasi *Cheese* → *Toothpaste* pada klaster ini mungkin tidak mencerminkan keterkaitan fungsional langsung, namun lebih menunjukkan kebiasaan berbelanja yang menyatukan berbagai kategori produk dalam satu transaksi yang merupakan ciri khas pelanggan

profesional. Untuk klaster ini, diskon besar bukanlah pendekatan utama. Sebaliknya, program loyalitas jangka panjang seperti sistem poin atau voucher belanja setelah akumulasi transaksi tertentu akan lebih efektif. Selain itu, pelanggan profesional juga cenderung menghargai efisiensi, sehingga promosi berbasis aplikasi atau notifikasi digital menjadi strategi yang tepat. Pendekatan ini selaras dengan pola konsumsi rasional dan loyalitas jangka panjang pelanggan profesional.

4.2. Perbandingan dengan Penelitian Terdahulu

Untuk memvalidasi hasil penelitian, diperlukan perbandingan dengan studi-studi terdahulu yang memiliki fokus serupa, khususnya pada penerapan algoritma FP-Growth maupun K-Means dalam analisis data transaksi ritel. Tabel 11 merupakan hasil penelitian terdahulu untuk dibandingkan dengan penelitian ini.

Tabel 11. Perbandingan dengan Penelitian Terdahulu

No	Penelitian	Dataset	Metode	Variabel	Hasil
1	[16]	Data transaksi retail dengan total 450 data	FP-Growth	Item pembelian per transaksi	Dihasilkan 4 aturan asosiasi dengan lift > 1 dan confidence ≥ 30%; digunakan untuk rekomendasi tata letak toko.
2	[17]	Data transaksi online retail dengan total 54910 data	K-Means (dengan Elbow Method)	<i>Quantity, Unit Price, Customer ID</i>	Terbentuk 4 klaster pelanggan, digunakan untuk strategi pemasaran.
3	[19]	Data transaksi pelanggan ritel dengan total 104347 data	RFM + K-Medoids, dilanjutkan FP-Growth	<i>Recency, Frequency, Monetary</i>	Terbentuk 5 klaster pelanggan, serta pola kombinasi produk untuk mendukung promosi.

Dari perbandingan penelitian terdahulu dalam tabel 11, dapat dilihat bahwa studi masih memisahkan proses segmentasi pelanggan dan analisis asosiasi produk. Penelitian ini memberikan kontribusi dengan mengintegrasikan algoritma K-Means dan FP-Growth secara langsung, sehingga pola belanja dapat dipetakan lebih spesifik pada setiap segmen pelanggan. Pendekatan ini menghasilkan strategi pemasaran yang lebih terarah sekaligus memperkuat literatur di bidang data mining dan sistem rekomendasi ritel.

5. KESIMPULAN

Penelitian ini menunjukkan bahwa kombinasi algoritma FP-Growth dan K-Means mampu memberikan pemahaman yang lebih mendalam terhadap perilaku pembelian pelanggan di industri ritel. Melalui analisis segmentasi pelanggan dengan K-Means, terbentuk tiga klaster pelanggan dengan karakteristik yang berbeda: *Retiree* yang cenderung responsif terhadap diskon langsung dan berbelanja produk fungsional, *Teenager* yang impulsif dan menyukai promosi BOGO, serta *Professional* yang minim ketergantungan pada diskon, namun memiliki nilai transaksi tinggi. Sementara itu, hasil analisis asosiasi dengan FP-Growth mengidentifikasi *Toothpaste* sebagai produk paling dominan dalam asosiasi pembelian lintas klaster, dengan pola seperti *Apple* → *Toothpaste* dan *Cleaning Spray* → *Toothpaste*.

Integrasi dari kedua metode ini menghasilkan strategi promosi yang lebih terarah, seperti bundling sederhana untuk pelanggan

retiree, hadiah langsung untuk pelanggan *teenager*, dan program loyalitas digital untuk pelanggan *professional*. Strategi ini dirancang tidak hanya berdasarkan data transaksi, tetapi juga mempertimbangkan kecenderungan psikologis dan preferensi tiap segmen. Meskipun penelitian ini masih terbatas pada atribut transaksi yang tersedia, hasilnya memberikan gambaran konkret bahwa pendekatan berbasis data mining dapat digunakan sebagai dasar perumusan rekomendasi produk dan strategi pemasaran yang lebih personal dan efektif. Untuk penelitian selanjutnya, disarankan untuk menambahkan variabel demografis atau perilaku lainnya guna memperkaya hasil segmentasi.

DAFTAR PUSTAKA

- Havid, “Ritel Modern Mengancam Keberadaan Toko Kelontong,” *Deras.id*. Internet: <https://deras.id/ritel-modern-mengancam-keberadaan-toko-kelontong/>, Mei 10, 2023 [Juli. 2, 2025].
- D. P. Pamelina, G. Sari, N. Faradea, dan J. H. K., “Pengaruh pasar modern terhadap pasar tradisional di era digital,” *Jurnal Review Pendidikan dan Pengajaran*, vol. 7, 2024.
- D. K. Chauhan, “Association Rule Mining for Market Basket Analysis on Online Retail Store,” *International Journal of Mechanical Engineering*, 2023, doi: [10.56452/6-3-671](https://doi.org/10.56452/6-3-671).
- Suherman, A. D. Arsita, dan I. Afriantoro, “Analisis Pola Penjualan di Toko Retail Menggunakan Algoritma Apriori dengan Rapid Miner,” *Jurnal Teknologi* <https://doi.org/10.25077/TEKNOSI.v11i2.2025.128-135>

- Informatika dan Komputer*, vol. 10, no. 2, hlm. 734–753, Des 2024, doi: [10.37012/jtik.v10i2.2407](https://doi.org/10.37012/jtik.v10i2.2407).
- [5] M. Aldyan Ruansyah, M. Zaky Hadi, M. Iqbal, J. Juniwati, dan L. Aulia, “Designing an analytics dashboard for knowledge extraction in the retail industry using descriptive and predictive analytics,” *Journal Industrial Servicess is*, vol. 10, no. 2, 2024, doi: [10.62870/jiss.v10i2.27900](https://doi.org/10.62870/jiss.v10i2.27900).
- [6] V. E. Putri dan H. D. Purnomo, “Integrasi algoritma apriori dan k-means dalam analisis pola pembelian untuk meningkatkan strategi pemasaran,” *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 1, hlm. 409–423, Jan 2025, doi: [10.29100/jipi.v10i1.5768](https://doi.org/10.29100/jipi.v10i1.5768).
- [7] A. Shinta dan K. Y. S. Putri, “Manajemen krisis public relations PT Hero supermarket pada penutupan gerai Giant Indonesia,” *Jurnal Ilmu Komunikasi*, 2022, [Daring]. Tersedia pada: <http://jurnal.utu.ac.id/jsource>
- [8] A. U. Khasanah dan M. R. Q. Baihaqie, “Analysis of consumer characteristics on retail business with clustering analysis method and association rule for selling improvement strategy recommendations,” *OPSI*, vol. 17, no. 1, hlm. 249, Jun 2024, doi: [10.31315/opsi.v17i1.11411](https://doi.org/10.31315/opsi.v17i1.11411).
- [9] A. Wijaya, A. Faqih, D. Solihudin, C. L. Rohmat, dan S. E. Permana, “Penerapan association rules menggunakan algoritma apriori untuk identifikasi pola pembelian,” *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 6, 2023.
- [10] F. Achmad, O. Nurdiawan, dan Y. A. Wijaya, “Analisa pola transaksi pembelian konsumen pada toko ritel kesehatan menggunakan algoritma FP-Growth,” *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 1, 2023, [Daring]. Tersedia pada: www.researchgate.net
- [11] M. Fathur Rezki Junaedi dan U. Hayati, “Analisis pola transaksi pembelian makanan dan minuman menggunakan algoritma FP-Growth,” *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 1, 2024.
- [12] F. N. Hasan dan R. Ariyansah, “Utilization of the FP-Growth Algorithm on MSME Transaction Data: Recommendations for Small Gifts from The Padang Region,” *Jurnal Teknik Informatika*, vol. 17, no. 1, hlm. 70–78, Mei 2024, doi: [10.15408/jti.v17i1.37966](https://doi.org/10.15408/jti.v17i1.37966).
- [13] S. L. Achmad, A. Fauzi, R. Rahmat, dan J. Indra, “Segmentasi pelanggan menggunakan k-means clustering di toko retail,” *Jurnal Teknik Informatika dan Komputer (Tekinkom)*, vol. 7, no. 2, hlm. 736, Des 2024, doi: [10.37600/tekinkom.v7i2.1226](https://doi.org/10.37600/tekinkom.v7i2.1226).
- [14] E. Prayitno, I. Jaka Perdana, E. Iskandar, B. H. Winarno, dan A. A. Subagyo, “Optimalisasi profitabilitas ritel melalui segmentasi pelanggan dengan k-means clustering,” *Jurnal Informasi Interaktif*, vol. 9, no. 3, 2024.
- [15] Zahrudin, A. Irma Purnamasari, dan I. Ali, “Analisis tren penjualan fashion import menggunakan algoritma fp-growth pada toko Air Gaul,” *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 3, no. 2, hlm. 75–84, 2024.
- [16] R. Dio dkk., “Analisis Data Mining Pembelian dengan Association Rule Market Basket Analysis menggunakan algoritma FP-Growth,” *Universitas Syiah Kuala Jl. Teuku Nyak Arief*, vol. XVII, no. 2, 2311.
- [17] E. Febrianty, L. Awalina, dan W. I. Rahayu, “Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail Optimizing Marketing Strategies with Customer Segmentation Using K-Means Clustering on Online Retail Transactions,” *Jurnal Teknologi dan Informasi (JATI)*, vol. 13, 2023, doi: [10.34010/jati.v13i2](https://doi.org/10.34010/jati.v13i2).
- [18] D. Putri, A. Asih, B. Irawan, dan A. Bahtiar, “Pengelompokan data transaksi dalam menentukan strategi penjualan menggunakan algoritma k-means,” *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 1, 2024.
- [19] T. A. Pertiwi, M. Afdal, R. Novita, dan M. Mustakim, “Penerapan Algoritma K-Medoids dan FP-Growth dengan Model RFM untuk Kombinasi Produk,” *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 2, hlm. 721–732, Sep 2024, doi: [10.47065/bits.v6i2.5268](https://doi.org/10.47065/bits.v6i2.5268).
- [20] Prasad Patil, “Retail Transactions Dataset,” Kaggle. Diakses: 10 Juli 2025. [Daring]. Tersedia pada: <https://www.kaggle.com/datasets/prasad22/retail-transactions-dataset/data>
- [21] A. Agung, A. Daniswara, I. Kadek, dan D. Nuryana, “Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru,” *Journal of Informatics and Computer Science*, vol. 05, 2023.
- [22] F. Dwi Insani, Kusriani, dan H. Al Fatta, “Perbandingan algoritma apriori dan fp-growth untuk rekomendasi item paket pada konten promosi di perusahaan Mu-Mart,” *Jurnal Bisnis Digital dan Sistem Informasi*, vol. 4, no. 2, hlm. 19–24, 2023.
- [23] I. Juwita dan I. Ali, “Penerapan pola penjualan dengan menggunakan metode algoritma asosiasi fp-growth bertujuan untuk meningkatkan penjualan kopi di Point Coffee,” *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 2, hlm. 1600, 2024.
- [24] F. A. W. T. S. Wijaya, E. Prasetyo, dan R. F. Tias, “Dynamic Time Warping Pada Metode K-Means Untuk Pengelompokan Data Trend Penjualan Produk,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 10, no. 2, hlm. 100–109, Agu 2024, doi: [10.25077/teknosi.v10i2.2024.100-109](https://doi.org/10.25077/teknosi.v10i2.2024.100-109).
- [25] I. Zuhdiansyah dan A. Luthfiarta, “Sistem Rekomendasi Pembelian Smartphone berbasis Algoritma K-Means dan Singular Value Decomposition,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 10, no. 1, hlm. 45–53, Mei 2024, doi: [10.25077/teknosi.v10i1.2024.45-53](https://doi.org/10.25077/teknosi.v10i1.2024.45-53).