Algoritma Apriori Dan K-Means Clustering Dengan GAP Rules Untuk Identifikasi Churn Dan Retensi Pelanggan

Woro Isti Rahayu[1)*], Fatia Amalia Maresti[2)*], Ahmad Mugiar Sujana[3)], Melvin Ariwati Hanek[4)]

[1,2,3,4]) Program Studi S1 Sains Data Universitas Logistik dan Bisnis Internasional

Received: 28 July 2025 Accepted: 26 December 2025 Published: 28 December 2025

*ahmadmugiars@gmail.com

Kata Kunci: K-Means Clustering, Market Basket Analysis, Analisis GAP Rules.

Abstrak: Peningkatan persaingan bisnis mendorong perusahaan untuk memahami perilaku pelanggan dengan lebih baik. Khususnya di bisnis retail, mempertahankan pelanggan adalah hal krusial yang harus diperhatikan. Oleh karena itu, sebagai pihak bisnis retail dapat memanfaatkan data transaksi untuk mengidentifikasi potensi *churn* dan menghasilkan strategi retensi berbasis data. Penelitian ini menggunakan pendekatan data mining melalui algoritma *K-Means Clustering* dengan karakteristik RFM*(Recency, Frequency, Monetary)* yang bertujuan untuk membuat segmentasi pelanggan, serta *Market Basket Analysis* menggunakan algoritma apriori. Hasil evaluasi *K-Means Clustering* menggunakan *Elbow Method* dan *Silhouette Score* sebesar 0.550, Menghasilkan 3 *Cluster* optimal. Dimana *cluster* 1 dilabelkan sebagai *Chrun Potential Customers*, *Cluster* 2 dilabelkan sebagai *Grow Potential Customers*, dan *Cluster* 3 sebagai *Loyal Customers*. Selanjutnya algoritma apriori ditetapkan minimum *support* sebesar 3% dan minimum *confidence* 30% menghasilkan beberapa aturan asosiasi yang memenuhi ambang batas yang ditetapkan. Hasilnya, *Cluster* 1 memiliki 3 aturan asosiasi Sedangkan *Cluster* 2 memiliki 13 aturan asosiasi. Dan *Cluster* 3 memiliki 15 aturan asosiasi yang memenuhi ambang batas. Perbedaan kekuatan asosiasi antar produk pada tiap *cluster* menunjukkan pola pembelian yang unik. Analisis GAP *Rules* di setiap *cluster* bertujuan untuk mengidentifikasi variasi pola pembelian barang yang dibeli bersamaan

PENDAHULUAN

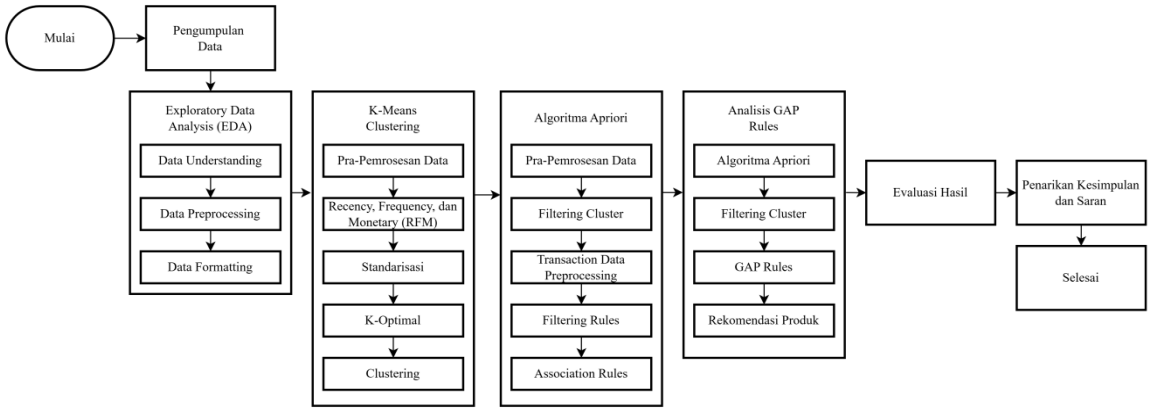
Pelanggan kini menjadi penggerak utama bisnis ritel di era digital. Kesuksesan dan profitabilitas bisnis secara langsung terkait dengan pemahaman perilaku pelanggan dan penyesuaian strategi pemasaran untuk memenuhi kebutuhan mereka[1]. Seiring meningkatnya persaingan dan pilihan konsumen, pelaku bisnis ritel tradisional maupun modern dituntut untuk memanfaatkan data dan informasi secara lebih efektif guna meningkatkan strategi pemasaran. Komponen penting dari pemasaran yang berhasil adalah segmentasi pelanggan, yaitu proses mengelompokkan pelanggan berdasarkan karakteristik yang sama[1]. Salah satu pendekatan untuk melakukan segmentasi pelanggan adalah *K-Means Clustering*. Pada penelitian ini, digunakan karakteristik pelanggan dari nilai RFM(*Recency, Frequency, Monetary*) sebagai input dalam *K-Means Clustering*. Pada penelitian ini, digunakan karakteristik pelanggan dari nilai RFM*(Recency, Frequency, Monetary)* sebagai input dalam *K-Means Clustering*. Model RFM untuk mengevaluasi perilaku pembelian pelanggan di tiga dimensi utama *recency*(Hitung berapa lama sejak transaksi terakhir yang dilakukan oleh setiap pelanggan hingga tanggal akhir periode analisis.), *frequency*(Hitung jumlah transaksi yang dilakukan oleh setiap pelanggan dalam periode analisis.), dan *monetary*(Hitung total nilai uang yang dibelanjakan oleh setiap pelanggan dalam periode analisis.)[2]. integrasi metode *Market Basket Analysis* menjadi langkah strategis, karena mampu mengungkap asosiasi antar produk yang sering dibeli secara bersamaan. Selain itu, diterapkan analisis GAP *rules* untuk membandingkan pola pembelian antar segmen, tujuannya adalah mengidentifikasi perbedaan aturan asosiasi pada setiap segmen yang didapatkan.

TINJAUAN LITERATUR

Beberapa peneliti sebelumnya telah mengkaji efektifitas metode *clustering* dengan *market basket analysis*. Penelitian oleh Sindy Genjang dkk (2021) dengan judul "*Market Basket Analysis with K-Means Clustering and FP-Growth as* Citra Mustika Pandawa *Company*"[3]. Penelitian tersebut menunjukan bahwa jumlah klaster optimal (K) adalah sebanyak 5, yang diperoleh melalui uji *validitas Davies Bouldin Indeks*(DBI) sebesar 0.500. Integrasi *K-Means Clustering* dan *FP-Growth* menghasilkan aturan asosiasi pada masing masing klaster, dengan dengan minimum *support* 30% dan minimum *confidence* 50%. Hasilnya, ditemukan 3 aturan asosiasi pada klaster 1, masing masing 1 aturan pada klaster 2 dan 3, kemudian 5 aturan asosiasi pada klaster 5, dan 6 aturan asosiasi pada klaster 5. Penelitian selanjutnya oleh Violita dkk (2025), dengan judul "Integrasi Algoritma Apriori dan *K-Means Clustering* dalam Analisis Pola Pembelian Untuk Meningkatkan Strategi Pemasaran"[4]. Penelitian pada UMKM Premium Salad.co ini membuktikan bahwa segmentasi pelanggan dan analisis pola pembelian dapat menjadi landasan strategi pemasaran yang efektif melalui pembuatan paket menu atau *bundling* produk. Dengan menetapkan ambang batas *minimum support* 0.01 dan *confidence* 0.5, ditemukan karakteristik transaksi yang berbeda pada setiap kelompok *Cluster* 0 (321 transaksi) menghasilkan 1 aturan asosiasi, *Cluster* 1 (228 transaksi) menghasilkan 3 aturan, dan *Cluster* 2 (127 transaksi) menjadi segmen paling potensial dengan 16 aturan serta tingkat kepercayaan mencapai 100%. Secara keseluruhan, hasil ini menunjukkan bahwa semakin spesifik klasternya, semakin kuat keterikatan antar produknya, sehingga perusahaan dapat menerapkan strategi *cross-selling* yang sangat akurat untuk meningkatkan frekuensi penjualan.

METODE PENELITIAN

Metode penelitian digunakan sebagai acuan agar proses penelitian dapat dilaksanakan secara sistematis, terarah, dan terstruktur. Dengan adanya metode yang tepat, pengumpulan serta pengolahan data dapat dilakukan secara lebih efektif guna mencapai tujuan yang telah ditetapkan.



Gambar 1. Metode Penelitian

Penelitian menggunakan pendekatan kuantitatif eksploratif dengan metode data mining [5], Metode yang digunakan dalam penelitian ini adalah integrasi antara segmentasi pelanggan berdasarkan karakteristik RFM(*Recency, Frequency, Monetary*) untuk membagi kelompok berdasarkan atribut yang digunakan, *Market Basket Analysis* menggunakan Algoritma Apriori untuk mengidentifikasi pola pembelian di setiap segmen yang didapatkan. Dan Analisis GAP *Rules* yang bertujuan untuk menemukan pola pembelian yang sering muncul pada segmen yang dijadikan sumber dengan segmen yang dijadikan target.

a. Pengumpulan Data
   Pengumpulan Data yang digunakan berasal dari dataset penjualan di sebuah toko retail periode transaksi dari tanggal 1 Januari 2022 hingga 31 Desember 2022.
b. *Exploratory Data Analysis*(EDA)
   Tahap EDA merupakan bagian krusial yang bertujuan untuk menggali dan memahami struktur data, mengidentifikasi pola, anomali, maupun potensi informasi penting yang tersembunyi di dalam data.
   1. Data *Understanding*
      Tahap data *understanding* meliputi memahami kolom kolom yang terdapat pada dataset.

2. Data *Preprocessing*
   Tujuan tahap ini mempersiapkan data secara menyeluruh, termasuk melakukan pembersihan data (data *cleaning*), sehingga data yang digunakan pada proses analisis selanjutnya sudah bersih, konsisten, dan siap diolah tanpa hambatan.
3. Data *Formatting*
   Pada tahap data *formatting*, diperlukan pengubahan tipe data pada kolom *transaction_date*, karena pada penelitian ini akan menghitung nilai *recency* pelanggan yang akan dilakukan pada tahap lanjutan.

c. *K-Means Clustering*
K-Means adalah algoritma pembelajaran mesin tanpa supervisi (*unsupervised learning*) yang digunakan untuk pengelompokan data (*clustering*), Algoritma ini bertujuan untuk membagi data ke dalam sejumlah kelompok (*cluster*) berdasarkan kemiripan atau kedekatan data dalam ruang fitur[6]. Segmentasi pasar menurut Philip Kotler dan Gary Amstrong adalah proses pembagian pasar menjadi segmen-segmen potensial berdasarkan kesamaan karakteristik yang mencerminkan perilaku pembeli yang serupa[7]. 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 [8].

1. Pra-Pemrosesan Data
   Tahap awal dalam proses segmentasi pelanggan adalah pra-pemrosesan data yang meliputi analisis *date references*. Dalam analisis RFM (*Recency, Frequency, Monetary*) khusus nya saat menentukan nilai *recency* pada setiap pelanggan, sangat penting untuk menetapkan tanggal referensi.
2. Model *Recency, Frequency, Monetary* (RFM)
   Model RFM (*Recency, Frequency, Monetary*) adalah sebuah model yang sudah banyak diimplementasikan di dalam dunia pemasaran karena dapat digunakan untuk mengambil keputusan secara efektif guna mengidentifikasi pelanggan yang berharga serta sebagai bahan pengembangan strategi pemasaran yang efektif [13]. Perhitungan nilai *Recency* dihitung dari (*reference date – last purchase date*) dalam hari, sedangkan *Frequency* yaitu *count*(*distinct transaction id*) dan yang terakhir adalah *Monetary* dihitung dari *sum*(*total*). Semua perhitungan nilai RFM dapat digabungkan berdasarkan *id customer*.
3. Standarisasi
   Standarisasi pada data RFM pelanggan, Tujuan dari standarisasi ini adalah untuk menyamakan skala antar variabel.
4. K-Optimal
   Pembentukan jumlah *cluster* terbaik berdasarkan dari beberapa cara salah satunya adalah *Elbow method* dan *Silhouette Score*.

$$WCSS = \sum_{k=1}^{k} \sum_{x_i \in S_k} ||x_i - c_k||^2$$

Keterangan :
k = *Cluster*
C_k = nilai rata-rata k *cluster*
x_i = data ke-i dalam *cluster*
$||x - c_k||^2$ = jarak kuadrat antara data dan *centroid*

Selain menggunakan *Elbow Method* penentuan K-Optimal juga di uji menggunakan *sillhouette Score*. *Silhouette Score* tidak jauh beda dengan *elbow method*, yakni untuk menentukan jumlah K yang optimal untuk proses pengelompokan lebih lanjut.

$$s(i) = \frac{b(i)-a(i)}{max(a(i),b(i))}$$

Keterangan :
$s(i)$ = *sillhouette score* untuk data ke-i
$a(i)$ = rata-rata jarak antar data ke-i dan semua anggota dalam *cluster*
$b(i)$ = rata-rata jarak antara data ke-i dan semua anggota terdekat

d. Algoritma Apriori

Algoritma apriori merupakan salah satu algoritma yang digunakan untuk *Market Basket Analysis*. *Market Basket Analysis* merupakan cabang dari ilmu data mining yang mempelajari pola pembelian barang oleh konsumen[9]. Untuk menemukan aturan-aturan asosiasi yang signifikan. Pendekatan ini dipilih karena kemampuannya dalam mengidentifikasi pola pembelian yang sering terjadi dan dapat memberikan wawasan yang dapat ditindaklanjuti untuk strategi bisnis [10].Tahap awal dalam proses ini adalah data transaksi di setiap segmen dikonversi menggunakan format data tabular, dengan membentuk data menggunakan konsep dari bilangan biner yaitu 0 dan 1. 1 berarti ada transaksi dan 0 tidak ada transaksi. Data tersebut diperoleh dari data rekapitulasi[11]. Cara ini menggunakan format data transaksional yang membutuhkan dua atribut, yaitu ID transaksi dan atribut konten atau isi belanja dari transaksi tersebut[12].

```
[ ]  # Encoding ke format 0-1
     te = TransactionEncoder()
     te_ary = te.fit(keranjang).transform(keranjang)
     basket_df = pd.DataFrame(te_ary, columns=te.columns_)

     basket_df
```

Gambar 2. TransactionEncoder

Hal ini dapat dilihat dari beberapa komponen nilai yang dihasilkan dari *Association rules* sebagai berikut :

1. *Support*

   Secara singkatnya nilai *support* adalah ukuran seberapa banyak barang A dibeli secara bersamaan dengan barang B. Adapun rumus perhitungan untuk mendapatkan nilai *support*.

$$Support\ (A) = \frac{\sum\ transaksi\ \ mengandung\ \ item\ \ A}{\sum\ total\ \ transaksi} . 100\%$$

2. *Confidence*

   Nilai kepercayaan(*Confidence*) adalah suatu nilai yang menunjukkan seberapa besar kemungkinan item B muncul di dalam transaksi. Berikut adalah rumus dari perhitungan nilai Kepercayaan (*Confidence*):

$$Confidence = \frac{Support\ (A \cap B)}{Support\ (A)} . 100\%$$

3. *Lift ration*

   Lift ratio dalam algoritma apriori adalah salah satu ukuran kekuatan asosiasi antara dua item atau itemset dalam *association rule*, berikut adalah perhitungan nilai *lift ratio*

$$Lift\ (A \rightarrow B) = \frac{Confidence\ (A \rightarrow B)}{Support\ (B)}$$

e. Analisis GAP *Rules* antar Segmen Pelanggan

GAP *Rules* yaitu aturan asosiasi yang muncul pada satu segmen namun tidak muncul pada segmen lain, sehingga dapat dimanfaatkan sebagai peluang rekomendasi produk. Pada tahap ini, aturan asosiasi dianggap identik (*matching*) apabila memiliki kesamaan *antecedent* dan *consequent* secara penuh (*exact match*). Aturan dinyatakan valid apabila memenuhi kriteria nilai *support*, *confidence*, dan *lift* > 1, sehingga GAP *Rules* yang diperoleh relevan secara statistik maupun bisnis dan dapat digunakan sebagai dasar strategi personalisasi promosi. Berikut adalah cara mencari GAP *rules*:

$$GAP(Source \rightarrow Target) = Rules\ Source - Rules\ Target$$

HASIL PENELITIAN

Pada bagian ini dibahas hasil integrasi antara algoritma *K-Means Clustering* dan algoritma Apriori. Kedua metode tersebut digunakan menganalisis data penjualan pada dataset yang digunakan. Evaluasi hasil meliputi hasil *clustering*, *association rules*, serta GAP *rules*. Dataset yang memiliki 5.020 baris data dan 18 kolom.

transaction_id	customer_id	transaction_date	product_i	price	qty	total_amo	store_id	age	gender	marital_s	income	product_name	store_name	group_store	type	latitude	longitude
T00001	C0328	2022-01-01 0:00:00	P3	7500	4	30000	12	36	0	Married	10,53	Crackers	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00001	C0328	2022-01-01 0:00:00	P9	10000	2	20000	12	36	0	Married	10,53	Yoghurt	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00001	C0328	2022-01-01 0:00:00	P1	8800	4	35200	12	36	0	Married	10,53	Choco Bar	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00002	C0017	2022-01-01 0:00:00	P9	10000	7	70000	12	44	1	Married	14,58	Yoghurt	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00002	C0017	2022-01-01 0:00:00	P9	10000	7	70000	12	44	1	Married	14,58	Yoghurt	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00003	C0215	2022-01-01 0:00:00	P9	10000	7	70000	12	44	1	Married	14,58	Yoghurt	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00003	C0215	2022-01-01 0:00:00	P1	8800	4	35200	12	27	1	Single	0,18	Choco Bar	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00003	C0215	2022-01-01 0:00:00	P1	8800	7	61600	12	48	1	Married	12,57	Choco Bar	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00004	C0050	2022-01-01 0:00:00	P9	10000	1	10000	12	33	0	Married	6,95	Yoghurt	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00004	C0050	2022-01-01 0:00:00	P10	15000	1	15000	12	19	1	Single	0	Cheese Stick	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00004	C0050	2022-01-01 0:00:00	P8	16000	2	32000	12	36	0	Married	7,95	Oat	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00005	C0454	2022-01-01 0:00:00	P5	4200	3	12600	12	44	1	Married	13,48	Thai Tea	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00006	C0041	2022-01-01 0:00:00	P9	10000	6	60000	12	45	0	Married	15,03	Yoghurt	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00006	C0041	2022-01-01 0:00:00	P7	9400	2	18800	12	49	1	Married	8,81	Coffee Candy	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00006	C0041	2022-01-01 0:00:00	P4	12000	4	48000	12	36	0	Single	3,7	Potato Chip	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00007	C0037	2022-01-01 0:00:00	P7	9400	2	18800	12	43	1	Married	5,69	Coffee Candy	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00007	C0037	2022-01-01 0:00:00	P2	3200	6	19200	12	34	0	Married	4,36	Ginger Candy	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00007	C0037	2022-01-02 0:00:00	P3	7500	4	30000	12	53	0	Married	17,2	Crackers	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00007	C0037	2022-01-02 0:00:00	P3	7500	2	15000	12	35	0	Married	9,01	Crackers	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00007	C0037	2022-01-02 0:00:00	P7	9400	4	37600	12	34	0	Married	10,96	Coffee Candy	Lingga	Lingga	Modern Trade	-5,1354	119,42379
T00008	C0138	2022-01-02 0:00:00	P10	15000	2	30000	12	53	1	Married	13,65	Cheese Stick	Lingga	Lingga	Modern Trade	-5,1354	119,42379

Gambar 3. Data Transaksi

1. *Exploratory Data Analysis*(EDA)

   Kolom yang akan digunakan adalah *transaction_id. customer_id, transaction_date, price, qty, total_amoun*t, dan *product_name*. Maka dari itu kolom yang tidak digunakan akan dihapus.

```
# Pilih kolom numerik dari df_cleaned
numerical_cols = ['price', 'qty', 'total_amount']

# Loop melalui setiap kolom numerik untuk mendeteksi outlier
for col in numerical_cols:
    Q1 = df_cleaned[col].quantile(0.25)
    Q3 = df_cleaned[col].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    outliers = df_cleaned[(df_cleaned[col] < lower_bound) | (df_cleaned[col] > upper_bound)]
    print(f"Outliers for column '{col}':")
    display(outliers)
    print("-" * 50)
```

Gambar 4. Analisis *Outlier*

Teknik kuartil digunakan mendeteksi *outlier* melalui Q1, Q2, dan Q3. Nilai yang berada jauh di luar rentang normal diidentifikasi sebagai *outlier*. Berdasarkan hasil identifikasi *outlier*, kolom *price* dan *total_amount* tidak memiliki *outlier*. kolom *qty* terdapat beberapa *outlier*, namun tidak dihapus karena masih masuk akal secara bisnis dan tidak mengandung nilai negatif.

Tabel 1. Dataset *Cleaning*

transaction_id	***customer_id***	***date***	***price***	***quantity***	***total***	***product***
T00001	C0328	01/01/2022	7500	4	30000	Crackers
T00001	C0328	01/01/2022	10000	2	20000	Yoghurt
T00001	C0328	01/01/2022	8800	4	35200	Choco Bar
T00002	C0017	01/01/2022	10000	7	70000	Yoghurt
T00002	C0017	01/01/2022	10000	7	70000	Yoghurt
T00003	C0215	01/01/2022	10000	7	70000	Yoghurt
T00003	C0215	01/01/2022	8800	4	35200	Choco Bar

Tabel diatas merupakan dataset atas yang akan digunakan dalam penelitian ini, jumlah baris sebanyak 5020 baris dan 7 kolom. Lalu pada dataset ini, produk unik nya itu ada crackers, yoghurt, choco bar, cheese stick, oat, thai tea, coffe candy, potato chip, ginger candy, dan cashew.

2. *K-Means Clustering*

   Terdapat 3 variabel yang akan digunakan untuk diproses dan diuji ke dalam analisis sesuai dengan atribut RFM yaitu *recency*, *frequency*, dan *monetary*.

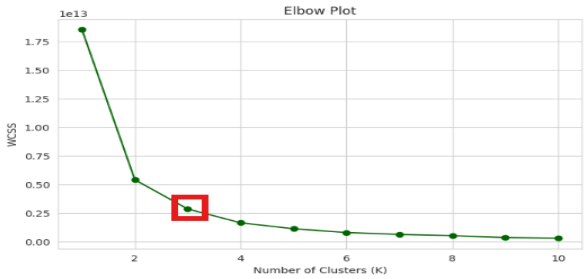
Tabel 2. *Recency, Frequency, Monetary Customers*

customer_id	***recency***	***frequency***	***monetary***
C0001	143	1	103.400,00
C0002	75	3	192.200,00
C0003	81	6	655.600,00
C0004	24	4	407.600,00
...	...	...	...

Langkah selanjutnya adalah standarisasi menggunakan *Z-Score* yang bertujuan untuk membuat nilai memiliki rata rata 0 dan *standard deviation* 1. Berikut adalah hasil dari standarisasi menggunakan perhitungan *Z-Score*.

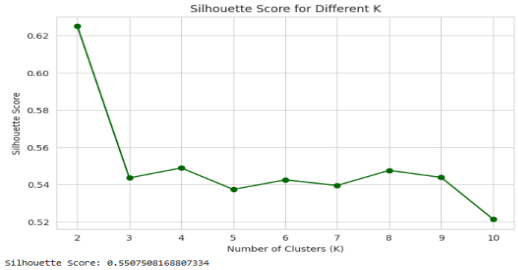
Tabel 3. Standarisasi Model RFM

customer_id	***recency***	***frequency***	***monetary***
C0001	0,512857069	-1,419837475	-1,166273674
C0002	-0,293636285	-0,245786991	-0,708734525
C0003	-0,222475107	1,515288734	1,678919095
...	...	...	...



Gambar 5. *Elbow Method*

Elbow Method terdapat nilai WCSS untuk beberapa nilai K, hasilnya diplot dalam grafik Titik siku (*elbow*) dari grafik menunjukkan K yang optimal[13]. Berdasarkan pengujian jumlah klaster menggunakan Elbow Method diperoleh titik siku (elbow) pada k = 3, yang menunjukkan bahwa pada titik tersebut penurunan nilai *Within Cluster Sum of Square* (WCSS) mulai melandai sehingga penambahan jumlah klaster setelah k = 3 tidak lagi memberikan peningkatan kualitas klaster signifikan. Selanjutnya pengujian menggunakan nilai *Silhouette Score*.



Gambar 6. *Silhouette Score*

Nilai *Silhouette Score* berkisar antara -1 hingga 1, di mana semakin mendekati 1, semakin baik klasterisasi[13]. Nilai *Silhouette Score* yang diperoleh adalah sebesar 0.550750, yang menunjukkan bahwa kualitas klaster yang terbentuk berada pada kategori cukup baik, di mana sebagian besar data telah terkelompokkan secara tepat, memiliki kohesi yang baik di dalam klaster, serta pemisahan antar klaster yang relatif jelas. Untuk lebih jelas lagi dalam pengambilan keputusan klaster terbaik, dapat dilakukan dengan uji validitas menggunakan *Davies Bouldin Index*(DBI).

Tabel 4. Uji Validitas DBI

Metode Evaluasi	**Cluster optimal**	**Nilai DBI**
Elbow Method	3	0.285585
Silhouette Score	3	0.550750

Suatu clusterdianggap optimal jika memiliki nilai DBI yang minimal[14]. Validasi ini diperkuat dengan hasil *Davies Bouldin Index* (DBI), di mana nilai DBI untuk k = 3 menunjukkan nilai yang relatif rendah dibandingkan k lainnya. Hal ini berarti klaster yang terbentuk memiliki tingkat kemiripan antar klaster yang rendah serta jarak antar klaster yang cukup baik (semakin kecil DBI menunjukkan klaster semakin baik). Dengan demikian, konsistensi hasil antara *Elbow Method* (WCSS), *Silhouette Score*, dan *Davies Bouldin Index* menunjukkan bahwa k = 3 merupakan jumlah klaster terbaik dan paling optimal

Tabel 5. Data Head Segmentasi Pelanggan

customer_id	***recency***	***frequency***	***monetary***	**label**
C0001	143	1	103.400	1
C0002	75	3	192.200	2

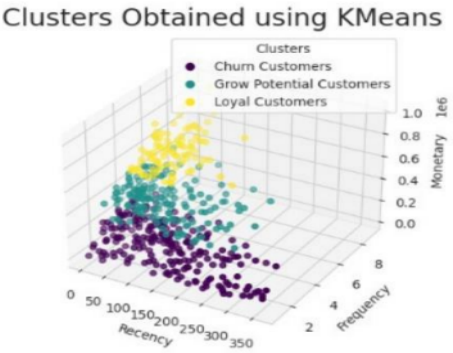
C0003	81	6	655.600
C0004	24	4	407.600
...	...	...	...

Tabel diatas menampilkan hasil perhitungan RFM untuk setiap pelanggan, kolom label menunjukan hasil pengelompokan(*cluter*) setiap pelanggan. Dengan adanya kolom label ini, setiap customer sudah memiliki identitas segmen masing masing sehingga bsa dianalisis karakteristiknya dan digunakan sebagai dasar penyusunan strategi. Menurut beberapa penelitian sebelumnya mendefisinikan churn berdasarkan periode ketidakatifan tertentu. Jika *recency* > 90 hari, maka pelanggan dikategorikan sebagai churn, dan Jika *recency* ≤ 90 hari, maka pelanggan dikategorikan sebagai *non-churn*[15]. Contoh pada baris pertama dengan *customer_id* C0001 dikategorikan sebagai *churn potential customers* karena nilai recency > 90 hari. Sedangkan *customer_id* C0002 hingga C0004 tidak di katakan sebagai *churn potential customers* karena nilai *recency* ≤ 90. Penelitian ini fokus pada identifikasi pelanggan *churn* & penyusunan strategi.

Tabel 5. Hasil *Clustering*

cluster	***recency***	***frequency***	***monetary***	**jumlah pelanggan**	**label**
1	133	2	163.745	224	*Churn Potential Customers*
2	82	3	373.347	169	*Grow Potential Customers*
3	52	5	633.206	99	*Loyal Customers*

Distribusi menunjukan bahwa mayoritas pelanggan berada dalam kategori potensial untuk tumbuh, sehingga dapat menjadi target utama untuk strategi pemasaran dan retensi yang lebih agresif guna meningkatkan loyalitas dan nilai pelanggan jangka panjang.



Gambar 7. Visualisasi Persebaran *Cluster*

Pada dimensi *recency*, semakin ke kanan titik data tersebar, maka nilai *recency* semakin tinggi, yang berarti pelanggan tersebut sudah cukup lama tidak melakukan transaksi (hingga 350 hari). Pada dimensi *Frequency*, titik yang berada semakin ke kanan atas menunjukkan frekuensi pembelian yang semakin tinggi. Sementara pada dimensi *monetary*, persebaran yang semakin ke atas menunjukkan bahwa pelanggan mengeluarkan nilai transaksi yang lebih besar.

Tabel 6. Kareakteristik *Cluster*

cluster	**kategori *Cluster***	**keterangan *Cluster***
1	*Potential Churn Customers*	Pelanggan dalam klaster ini sudah lama tidak bertransaksi dan menunjukkan keterlibatan rendah.
2	*Grow Potential Customers*	Pelanggan ini memiliki nilia *recency, frequency,* dan *monetary* ditingkat menengah
3	*Loyal Customers*	Pelanggan dalam klaster ini ditandai dengan rata rata *recency* yang rendah, serta nilai *freuency* dan *monetary* tertinggi dibandingkan klaster lainnya.

3. Algoritma Apriori

Algoritma Apriori adalah jenis Aturan Asosiasi dalam penambangan data. Aturan yang menyatakan asosiasi antara atribut sering disebut analisis afinitas atau analisis keranjang pasar. Analisisasosiasi atau *association rule mining* adalah teknik data mining untuk menemukan aturan dari kombinasi item[16]. Dalam penelitian ini,nilai ambang batas tersebut akan dijadikan acuan utama dalam proses pembentukan *Association Rules* yang valid dan relevan terhadap pola pembelian pelanggan.

Tabel 7. Ambang Batas Nilai Aturan Asosiasi

Minimum Support	*Minimum Confidence*	*Minimum Lift Ratio*
3%	30%	1

Minimal support nilai minimal dari kombinasi itemset[16]. Nilai *confidence* adalah ukuran probabilitas bahwa item B akan dibeli jika item A sudah dibeli[17]. Uji *lift ratio* ini digunakan untuk menguji kekuatan *rule* yang sudah terbentuk[18].

a. *Association Rules Cluster 1*

Tabel 8. Aturan Asosiasi *Cluster* 1

No	Rule	Support	Confidence	Lift
1	cashew --> cheese stick	3.0%	31.30%	1.177326
2	potato chip --> thai tea	7,4%	42,85%	1.041429
3	oat --> thai tea	9,2%	41,12%	1.003211

Ada 3 aturan asosiasi dihasilkan dari *cluster* 1 dan telah memenuhi ambang batas minimum, menunjukkan adanya pola pembelian yang signifikan di kelompok pelanggan ini. Aturan dengan nilai lift tertinggi adalah "jika produk potato chip dibeli, maka thai tea juga akan dibeli", dengan nilai *support* sebesar 7.4%, *confidence* 42.85%, dan *lift* sebesar 1.041429.

b. *Association Rules Cluster 2*

Tabel 9. Aturan Asosiasi *Cluster 2*

No	*Rule*	*Support*	*Confidence*	*Lift*
1.	coffee candy, crackers --> cheese stick	3.47%	45.84%	1.397035
2.	coffee candy, thai tea --> cheese stick	3.94%	40.32%	1.229063
3.	cheese stick, crackers --> coffee candy	3.47%	34.37%	1.204075
..	.....	.....	.....	.....
13	potato chip --> thai tea	7.88%	38.59%	1.019653

Dalam *cluster* 2, d*item*ukan 13 aturan asosiasi yang memiliki nilai *support*, *confidence*, dan *lift* sesuai dengan kriteria minimum, sehingga dianggap layak untuk dijadikan dasar pengambilan keputusan dalam analisis pembelian. Aturan dengan nilai lift tertinggi adalah "jika produk potato chip dibeli, maka produk thai tea juga akan dibeli", dengan nilai *support* sebesar 7.88% , *confidence* 38.59%, dan *lift* sebesar 1. 019653.

c. *Association Rules Cluster 3*

Tabel 10. Aturan Asosiasi *Cluster 3*

No	*Rules*	*Support*	*Confidence*	*Lift Ratio*
1	coffee candy, thai tea --> ginger candy	3.20%	37.50%	1.359677
2	potato chip, thai tea --> cheese stick	3.38%	45.23%	1.258604
3	coffee candy, thai tea --> cheese stick	3.58%	41.66%	1.159241
..	.....	.....	.....	.....
15	thai tea --> cheese stick	14.99%	41.15%	1.062286

Berdasarkan hasil pemrosesan data pada *cluster* 3, sistem menghasilkan 15 aturan asosiasi yang lolos penyaringan menggunakan ambang minimum *support*, *confidence*, dan lift yang telah ditentukan sebelumnya. Aturan dengan nilai lift tertinggi adalah "jika produk thai tea dibeli, maka produk cheese stick juga akan dibeli", dengan nilai *support* sebesar 14.99%, *confidence* 41.15%, dan *lift* sebesar 1.062286. Aturan ini dapat dimanfaatkan sebagai rekomendasi produk secara personal untuk pelanggan di *cluster* 3.

3. Analisis GAP *Rules*

Analisis GAP Rules mengidentifikasi perbedaan atau kesenjangan aturan asosiasi antar segmen pelanggan, khususnya dalam melihat produk-produk yang memiliki potensi untuk direkomendasikan namun belum muncul sebagai kombinasi pembelian yang dominan di suatu *cluster*.

```
def rules_to_set(rules_df):
    return set((frozenset(a), frozenset(c)) for a, c in zip(rules_df['antecedents'], rules_df['consequents']))
set_a = rules_to_set(top10_rules_a)
set_b = rules_to_set(top10_rules_b)
gap_rules = set_a - set_b
```

Gambar 8. *Frozenset Rules*

Diterapkan pada aturan asosiasi *cluster source* dan *cluster* target target untuk membuat *set* a dan *set* b. Dimana setiap aturan direpresentasikan sebagai pasangan *frozenset* dari *antecedent* dan *consequent*. lalu menghitung selisih antara kedua set tersebut yang menghasilkan GAP *Rules*.

Tabel 11. Hasil GAP *Rules*

***Cluster* Source**	**Aturan Asosiasi *Source***	***Cluster* Target**	**Aturan Asosiasi Target**	**GAP *Rules***
Cluster 3	15 *Rules*	*Cluster* 1	3 *Rules*	15 GAP *Rules*
Cluster 3	15 *Rules*	*Cluster* 2	13 *Rules*	13 GAP *Rules*
Cluster 2	13 *Rules*	*Cluster* 1	3 *Rules*	13 GAP *Rules*

Analisis GAP *Rules* mengungkap adanya kesenjangan pola pembelian antar-*cluster* yang dapat dimanfaatkan sebagai dasar rekomendasi produk. Pada hubungan *Cluster* 3 (*Source*) terhadap *Cluster* 1 (Target), diperoleh 15 aturan asosiasi, namun tidak ditemukan aturan yang memiliki kesamaan pola pembelian (*overlap*) dengan *Cluster* 1. Artinya, seluruh 15 *rules* tersebut sepenuhnya merupakan potensi pengetahuan dari *Cluster* 3 yang dapat direkomendasikan kepada Cluster 1 sebagai kandidat strategi peningkatan aktivitas transaksi. Selanjutnya, sebanyak 15 aturan dari *Cluster* 3 diarahkan ke *Cluster* 2, dan 13 aturan dari *Cluster* 2 diarahkan ke *Cluster* 1.

PEMBAHASAN

Algoritma Apriori adalah untuk mengidentifikasi *association rules*, yaitu aturan asosiasi yang menggambarkan keterkaitan antar produk yang sering dibeli bersamaan dalam satu transaksi. Melalui penerapan metode ini pada masing-masing *cluster*, diperoleh wawasan yang lebih spesifik terkait kebiasaan belanja dari setiap segmen pelanggan.

Tabel 12. Implikasi Strategi

Cluster	***Top Association Rule***	**Nilai Utama**			**Implikasi Strategi**
		Support	***Confidence***	***Lift***	
Cluster 1	Potato Chip → Thai Tea	7.40%	42.85%	1.041	*Promo bundling* Potato Chip dan Thai Tea untuk meningkatkan nilai transaksi
Cluster 2	Potato Chip → Thai Tea	7.88%	38.59%	1.020	*Upselling* & rekomendasi otomatis kombinasi produk saat pembelian
Cluster 3	Thai Tea → Cheese Stick	14.99%	41.15%	1.062	*Cross-selling* agresif melalui paket Thai Tea + Cheese Stick & loyalty promotion

KESIMPULAN

Penelitian mengidentifikasi 3 *cluster* pelanggan menggunakan model RFM. *cluster* 1 sebanyak 224 pelanggan dikategorikan sebagai *churn potential customers*, *cluster* 2 sebanyak 169 pelanggan dikategorikan sebagai *grow potential customers*, dan *cluster* 3 sebanyak 99 pelanggan dikategorikan sebagai *loyal customers*. Dalam algoritma apriori menghasilkan 3 aturan di *cluster* 1, 13 aturan di *cluster* 2, dan 15 aturan di *cluster* 3. tahapan terakhir adalan analisis GAP *Rules* menghasilkan sebanyak 15 aturan dari *Cluster* 3 direkomendasikan untuk *Cluster* 1 (dapat menjadi dasar dalam merancang intervensi untuk meningkatkan keterlibatan), 13 aturan dari *Cluster* 3 direkomendasikan untuk *Cluster* 2 (untuk retensi dan promosi bertarget), 13 aturan dari *Cluster* 2 direkomendasikan untuk *Cluster* 1 (memberikan rekomendasi yang dapat dimanfaatkan dalam upaya retensi dan pengembangan pelanggan). Penelitian ini menunjukkan bahwa produk yang lebih populer dan lebih aktif dibeli pada cluster sumber dapat dijadikan rekomendasi bagi *cluster* target yang aktivitasnya lebih rendah. Secara keseluruhan, Thai Tea adalah "*Anchor Product*" (produk penjangkar). Strategi terbaik adalah memposisikan Thai Tea sebagai pusat promosi, kemudian menyandingkannya dengan *snack* yang relevan berdasarkan preferensi *cluster* masing-masing untuk meningkatkan *Average Transaction Value* (ATV) Dengan demikian, GAP *Rules* berperan dalam mendukung strategi pemasaran yang lebih efektif, seperti *cross-selling* dan *bundling product*, sehingga mampu mengisi celah kebutuhan, meningkatkan relevansi penawaran, serta memperbesar peluang peningkatan transaksi pada segmen pelanggan dengan tingkat aktivitas lebih rendah.

REREFENCES

[1] A. Yusak, N. Rumapea, D. Pratiwi, and S. Sari, "Analisis Segmentasi Pelanggan Ritel Online Menggunakan K-Means Clustering Berdasarkan Model Recency, Frequency, Monetary (RFM)," *Jurnal Sains dan Teknologi*, vol. 6, no. 3, pp. 292–299, 2024.

[2] Y. Michael Sihombing, P. yuspandi, and J. Rimaza putra, "Segmentasi Pelanggan Menggunakan Fuzzy C-Means Clustering berdasarkan RFM Model pada E-Commerce ABC," *JRIIN: Jurnal Riset Informatika*

dan Inovasi, vol. 2, no. 1, 2024, [Online]. Available: https://jurnalmahasiswa.com/index.php/jriin

[3] S. G. Setyorini, E. K. Sari, L. R. Elita, and S. A. Putri, "Analisis Keranjang Pasar Menggunakan Algoritma K-Means dan FP-Growth pada PT. Citra Mustika Pandawa," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 41–46, 2021, doi: 10.57152/malcom.v1i1.62.

[4] V. E. Putri and 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, pp. 409–423, 2025, doi: 10.29100/jipi.v10i1.5768.

[5] H. Pratama, P. Nabawy, B. Cahyadi, and M. Furqan, "Analisis Pola Asosiasi Interaksi Pengguna pada Sistem Informasi Akademik Berbasis Web Menggunakan Algoritma Apriori," vol. 5, no. 1, pp. 10–17, 2025.

[6] N. Septiani and S. Wahyuni, "Bulletin of Information Technology (BIT) Implementasi Data Mining Dalam Mengelompokkan Tingkat Kepuasan Pemakaian Jasa Cleaning Service Dengan Menggunakan Algoritma K-Means Clustering," vol. 5, no. 4, pp. 340–354, 2024, doi: 10.47065/bit.v5i2.1729.

[7] A. H. Faza *et al.*, "ANALISIS SEGMENTASI PASAR DALAM PERENCANAAN BISNIS INDUSTRI RITEL," *Jurnal Ilmiah Multidisiplin*, vol. 1, no. 4, pp. 40–49, 2024, doi: 10.62017/merdeka.

[8] G. B. Sulistyo, N. Hasan, S. Kiswati, F. Natalia, E. Muningsih, and P. Korespondensi, "Segmentasi Pelanggan dan Optimalisasi Penjualan pada Data Retail Online Berbasis Model RFM," *Computer and Network Technology*, vol. 5, no. 1, pp. 16–22, 2025, [Online]. Available: http://jurnal.bsi.ac.id/index.php/conten16

[9] W. D. Ramadana, N. Satyahadewi, and H. Perdana, "Penerapan Market Basket Analysis Pada Pola Pembelian Barang Oleh Konsumen Menggunakan Metode Algoritma Apriori," *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, vol. 11, no. 3, pp. 431–438, 2022.

[10] Hendri, "ANALISIS POLA TRANSAKSI PENGGUNA MENGGUNAKAN ALGORITMA ASOSIASI PADA DATA E-COMMERCE Hendri 1) 1)," vol. 01, Jul. 2025, Accessed: Dec. 22, 2025. [Online]. Available: https://sihojurnal.com/index.php/jitifna/article/view/744/538

[11] F. Nuryawan and E. Mailoa, "ANALISIS POLA MINAT KONSUMEN DENGAN ALGORITMA APRIORI ARTICLE INFO," vol. 15, no. 2, pp. 269–276, 2024, [Online]. Available: http://ejurnal.provisi.ac.id/index.php/JTIKP

[12] Muhammad Rafly Qowi Baihaqie, "ANALISIS PERILAKU KONSUMEN PADA USAHA RITEL DENGAN MENGGUNAKAN METODE ASSOCIATION RULE - MARKET BASKET ANALYSIS DAN CLUSTERING SEBAGAI USULAN STRATEGI PENINGKATAN PENJUALAN (Studi Kasus: INTIMART GEDONGAN)," *Aleph*, vol. 87, no. 1,2, pp. 149–200, 2023.

[13] A. F. Afra, A. L. Hananto, A. Hananto, and B. Priyatna, "Klasterisasi Supplier Berdasarkan Kinerja Menggunakan Algoritma K-Means," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 7, no. 2, pp. 334–341, 2025, doi: 10.47233/jteksis.v7i2.1935.

[14] Muhammad Raqib Syahkur, D. Hartama, and S. Solikhun, "Evaluasi Jumlah Cluster pada Algoritma K-Means++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita," *JST (Jurnal Sains dan Teknologi)*, vol. 13, no. 3, pp. 487–496, Oct. 2024, doi: 10.23887/jstundiksha.v13i3.86419.

[15] S. D. Yuliani and P. Kartikasari, "KLASIFIKASI PELANGGAN CHURN PADA KEGIATAN SERVIS DAN PENJUALAN SPAREPART DENGAN METODE RANDOM FOREST," *Prosiding Seminar Nasional Sains dan Teknologi Seri IV Fakultas Sains dan Teknologi*, vol. 2, no. 2, 2025.

[16] B. Dwi Meilani and A. Kharis Juniawan, "Menentukan Pola Materi Sulit Menggunakan Association Rule Mining," 2025. [Online]. Available: https://edu.pubmedia.id/index.php/jtp

[17] R. Junianto and H. M. Nawawi, "Analisis Pola Penjualan pada Coffee Shop Menggunakan Algoritma Apriori (Studi Kasus: Kopislashtea)," *Jurnal Teknologi dan Informasi*, vol. 15, no. 1, pp. 29–39, Mar. 2025, doi: 10.34010/jati.v15i1.14229.

[18] D. A. Valeska, F. R. Umbara, and P. N. Sabrina, "Diagnosa Gejala yang Muncul Bersamaan pada Penderita Tuberculosis Menggunakan Algoritma Apriori dengan Substitusi Metode Bayesian pada Nilai Confidence," *Jurnal Teknologi Informatika dan Komputer*, vol. 8, no. 1, pp. 318–332, 2022, doi: 10.37012/jtik.v8i1.1105.