

Customer Intelligence Berbasis RFM Clustering dan Market Basket Analysis untuk Optimalisasi Strategi Bisnis pada Online Retail

Abdul Rasyid¹, Felix Cris Rino², Daffa Alifya Rayhan³

^{1,2,3}Program Studi Sistem Informasi, Universitas Pamulang

abdulrasyid.hadi02@gmail.com¹, felixpros182@gmail.com², alifyarayhan@gmail.com³

Article Info

Article history:

Received April 25, 2026

Revised May 18, 2026

Accepted June 20, 2026

Keywords:

Customer Intelligence, RFM, K-Means Clustering, Market Basket Analysis, FP-Growth, Online Retail

ABSTRACT

Competition in the digital retail industry is intensifying, making a deep understanding of customer behavior crucial for business success. This study implements a Customer Intelligence approach by integrating two data mining techniques—K-Means Clustering based on RFM (Recency, Frequency, Monetary) values and Market Basket Analysis (MBA) using the FP-Growth algorithm—on the Online Retail II dataset from the UCI Machine Learning Repository, covering 1,067,371 transactions from December 2009 to December 2011. After data cleaning, 779,425 clean rows (73.0% of original data) and 5,878 unique customers were obtained. RFM values were computed using the reference date of December 10, 2011, yielding an average Recency of 201.33 days, Frequency of 6.29 transactions, and Monetary of £2,955.90. The Silhouette Score method produced the highest score of 0.4386 at K=2. Two customer segments emerged: Loyal Customers (2,320 customers, 39.5%; Avg R=50.56 days, F=12.72x, M=£6,546.97) and At-Risk/Lost (3,558 customers, 60.5%; Avg R=299.65 days, F=2.10x, M=£614.34). FP-Growth analysis with a minimum support of 2% produced 180 frequent itemsets and 15 association rules with lift greater than 1. The highest-lift rule (13.95) was found between STRAWBERRY CERAMIC TRINKET BOX and SWEETHEART CERAMIC TRINKET BOX (support 2.25%, confidence 45.62%). Total dataset revenue reached £17,374,804. Based on these findings, strategic recommendations including loyalty programs, win-back campaigns, and product bundling were formulated to enhance customer retention and average transaction value.

This is an open access article under the CC BY-SA license.



Article Info

Article history:

Received April 25, 2026

Revised May 18, 2026

Accepted June 20, 2026

Keywords:

Intelijen Pelanggan, RFM, K-Means Clustering, Analisis Keranjang Pasar, FP-Growth, Ritel Online

ABSTRAK

Persaingan dalam industri ritel digital semakin ketat, sehingga pemahaman mendalam terhadap perilaku pelanggan menjadi krusial untuk keberhasilan bisnis. Penelitian ini menerapkan pendekatan Customer Intelligence dengan mengintegrasikan dua teknik data mining, yaitu K-Means Clustering berbasis nilai RFM (Recency, Frequency, Monetary) dan Market Basket Analysis (MBA) menggunakan algoritma FP-Growth, pada dataset Online Retail II dari UCI Machine Learning Repository yang mencakup 1.067.371 transaksi selama periode Desember 2009 hingga Desember 2011. Setelah melalui proses pembersihan data, diperoleh 779.425 baris data bersih (73,0% dari data asli) dan 5.878 pelanggan unik. Nilai RFM dihitung menggunakan tanggal referensi 10 Desember 2011 dengan hasil: rata-rata Recency 201,33 hari, rata-rata Frequency 6,29 transaksi, dan rata-rata Monetary £2.955,90. Metode Silhouette Score menghasilkan nilai tertinggi K=2 sebesar 0,4386. Dua segmen pelanggan yang terbentuk adalah Loyal Customers (2.320 pelanggan, 39,5%; Avg R=50,56 hari, F=12,72x, M=£6.546,97) dan At-Risk/Lost (3.558 pelanggan, 60,5%; Avg R=299,65 hari, F=2,10x, M=£614,34). Analisis FP-Growth dengan minimum support 2% menghasilkan 180 frequent itemsets dan 15 association rules dengan lift lebih dari 1. Aturan dengan lift tertinggi (13,95) ditemukan antara

STRAWBERRY CERAMIC TRINKET BOX dan SWEETHEART CERAMIC TRINKET BOX (support 2,25%, confidence 45,62%). Total revenue dataset mencapai £17.374.804. Berdasarkan temuan tersebut dirumuskan rekomendasi strategis berupa program loyalitas, win-back campaign, dan bundling produk untuk meningkatkan retensi dan nilai transaksi pelanggan.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Abdul Rasyid
Universitas Pamulang
Email: abdulrasyid.hadi02@gmail.com

PENDAHULUAN

Industri retail online mengalami pertumbuhan yang signifikan dalam satu dekade terakhir. Menurut Statista (2024), nilai pasar e-commerce global diperkirakan mencapai USD 6,3 triliun pada tahun 2024. Di tengah pertumbuhan ini, pelaku bisnis retail dihadapkan pada tantangan untuk memahami perilaku pelanggan secara lebih mendalam guna mempertahankan daya saing. Dataset Online Retail II dari UCI Machine Learning Repository merekam 1.067.371 transaksi dari sebuah perusahaan retail online berbasis di United Kingdom selama periode Desember 2009 hingga Desember 2011, menjadi bukti nyata kompleksitas dan volume data yang harus dikelola pelaku bisnis modern.

Customer Intelligence merupakan proses pengumpulan dan analisis data pelanggan untuk memperoleh wawasan yang dapat digunakan dalam pengambilan keputusan bisnis (Kumar & Reinartz, 2018). Pendekatan ini semakin relevan seiring meningkatnya ketersediaan data transaksi yang dapat dianalisis menggunakan teknik data mining. Dua teknik yang umum digunakan dalam konteks Customer Intelligence adalah segmentasi pelanggan berbasis nilai RFM dan analisis asosiasi produk.

Analisis RFM (Recency, Frequency, Monetary) mengklasifikasikan pelanggan berdasarkan tiga dimensi perilaku pembelian: seberapa baru pelanggan melakukan transaksi, seberapa sering, dan berapa besar nilai transaksi yang dihasilkan (Hughes, 1994). Metode ini diintegrasikan dengan algoritma K-Means Clustering untuk menghasilkan segmen pelanggan yang dapat diimplementasikan secara praktis. Di sisi lain, Market Basket Analysis (MBA) menggunakan algoritma FP-Growth memungkinkan identifikasi pola pembelian produk secara bersamaan yang dapat dimanfaatkan untuk strategi cross-selling dan bundling produk.

Pertanyaan penelitian yang dirumuskan adalah: (1) Bagaimana segmentasi pelanggan berdasarkan nilai RFM dengan K-Means Clustering dapat mengidentifikasi profil pelanggan yang berbeda untuk mendukung strategi retensi? (2) Pola pembelian produk apa yang dapat diidentifikasi melalui Market Basket Analysis untuk mendukung strategi bundling dan cross-selling? Tujuan akhir penelitian ini adalah menghasilkan rekomendasi bisnis yang actionable berdasarkan temuan dari kedua teknik data mining tersebut.

Tinjauan Pustaka

Customer Intelligence dan Data Mining

Customer Intelligence (CI) didefinisikan sebagai proses pengumpulan dan analisis informasi tentang pelanggan dan pasar dalam rangka mendapatkan wawasan mendalam yang digunakan untuk pengambilan keputusan bisnis yang lebih baik (Buttle & Maklan, 2019). Dalam era digital, CI semakin bergantung pada teknik-teknik data mining yang mampu mengolah volume data besar secara efisien.

Data mining merupakan proses penemuan pola dan pengetahuan yang berguna dari kumpulan data besar (Han et al., 2011). Teknik-teknik dalam data mining meliputi clustering, klasifikasi, regresi, dan analisis asosiasi. Dalam konteks retail, teknik clustering dan analisis asosiasi paling sering diaplikasikan untuk memahami segmentasi pelanggan dan pola pembelian produk.

Analisis RFM

Model RFM pertama kali diperkenalkan oleh Hughes (1994) dan telah menjadi salah satu kerangka analisis pelanggan yang paling banyak digunakan dalam industri retail. RFM terdiri dari tiga dimensi utama: Recency (R) mengukur seberapa baru seorang pelanggan melakukan transaksi terakhir, Frequency (F) mengukur seberapa sering pelanggan melakukan pembelian dalam periode tertentu, dan Monetary (M) mengukur total nilai pembelian pelanggan. Secara matematis, ketiga dimensi tersebut dirumuskan sebagai berikut:

$$\text{Recency} = \text{Tanggal Referensi} - \text{Tanggal Transaksi Terakhir}$$

Persamaan (1) — Recency

$$\text{Frequency} = \text{COUNT_UNIQUE}(\text{Invoice})$$

Persamaan (2) — Frequency

$$\text{Monetary} = \sum_{i=1}^n \text{TotalPrice}_i$$

Persamaan (3) — Monetary

Keterangan:

TotalPrice = Quantity × Price untuk setiap baris transaksi

n = jumlah total transaksi (baris) milik pelanggan tersebut

Penelitian Tsai & Chiu (2004) menunjukkan bahwa model RFM efektif dalam mengidentifikasi pelanggan bernilai tinggi dengan akurasi yang lebih baik dibandingkan metode statistik konvensional. Selanjutnya, Wei et al. (2010) mengintegrasikan nilai RFM dengan algoritma K-Means untuk menghasilkan segmentasi pelanggan yang lebih akurat dan bermakna secara bisnis.

K-Means Clustering

K-Means adalah algoritma clustering yang paling banyak digunakan karena kesederhanaan dan efisiensinya (MacQueen, 1967). Algoritma ini bekerja dengan cara mengalokasikan setiap titik data ke cluster terdekat berdasarkan jarak Euclidean terhadap centroid cluster, yang dirumuskan sebagai berikut:

$$d(\mathbf{x}, \mathbf{c}) = \sqrt{\sum_{i=1}^k (\mathbf{x}_i - \mathbf{c}_i)^2}$$

Persamaan (4) — Jarak Euclidean antara titik data dan centroid

Keterangan:

x = vektor fitur titik data (R_scaled, F_scaled, M_scaled)

c = vektor centroid cluster

k = jumlah dimensi fitur (k = 3 untuk RFM)

Iterasi dilakukan hingga centroid tidak berubah secara signifikan, dengan tujuan meminimalkan total Within-Cluster Sum of Squares (Inertia), yang dirumuskan sebagai berikut:

$$\text{Inertia} = \sum_{j=1}^K \sum_{\mathbf{x} \in C_j} \|\mathbf{x} - \mathbf{c}_j\|^2$$

Persamaan (5) — Inertia (Within-Cluster Sum of Squares)

Keterangan:

K = jumlah cluster

C_j = himpunan titik data pada cluster ke-j

c_j = centroid cluster ke-j

Penentuan jumlah cluster yang optimal umumnya dilakukan menggunakan Elbow Method yang memplot nilai inertia terhadap jumlah cluster, dan Silhouette Score yang mengukur kualitas cluster berdasarkan kohesi dan separasi (Rousseeuw, 1987). Silhouette Score untuk satu titik data dihitung dengan rumus:

$$s(i) = [b(i) - a(i)] / \max\{a(i), b(i)\}$$

Persamaan (6) — Silhouette Score untuk satu titik data ke-*i*

Keterangan:

$a(i)$ = rata-rata jarak titik *i* terhadap seluruh titik lain dalam cluster yang sama (kohesi)

$b(i)$ = rata-rata jarak titik *i* terhadap seluruh titik pada cluster terdekat lainnya (separasi)

$s(i)$ = nilai Silhouette Score titik *i*, berkisar antara -1 hingga 1

Nilai Silhouette Score keseluruhan merupakan rata-rata $s(i)$ untuk seluruh titik data dalam dataset. Nilai mendekati 1 menunjukkan kualitas cluster yang baik (titik berada jauh dari cluster tetangga), nilai mendekati 0 menunjukkan cluster yang saling tumpang tindih, dan nilai negatif menunjukkan kesalahan penempatan cluster.

Market Basket Analysis dan FP-Growth

Market Basket Analysis (MBA) bertujuan menemukan pola produk yang sering dibeli secara bersamaan dalam satu transaksi. Teknik ini menggunakan konsep association rules dengan tiga metrik utama (Agrawal & Srikant, 1994). Untuk sebuah aturan asosiasi $A \rightarrow B$ (jika membeli produk A, maka membeli produk B), ketiga metrik tersebut dirumuskan sebagai berikut:

$$\text{Support}(A \rightarrow B) = \text{Jumlah transaksi mengandung A dan B} / \text{Total transaksi}$$

Persamaan (7) — Support

$$\text{Confidence}(A \rightarrow B) = \text{Support}(A \rightarrow B) / \text{Support}(A)$$

Persamaan (8) — Confidence

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

Persamaan (9) — Lift

Keterangan:

$\text{Support}(A)$ = proporsi transaksi yang mengandung item A terhadap seluruh transaksi

$\text{Confidence}(A \rightarrow B)$ = probabilitas bersyarat membeli B jika telah membeli A

$\text{Lift}(A \rightarrow B)$ = kekuatan asosiasi A dan B dibandingkan jika keduanya independen; Lift = 1 berarti tidak ada asosiasi, Lift > 1 berarti asosiasi positif

Algoritma FP-Growth (Frequent Pattern Growth) yang diperkenalkan oleh Han et al. (2000) merupakan alternatif yang lebih efisien dibandingkan Apriori karena menghindari pembangkitan candidate itemset yang berjumlah besar. FP-Growth menggunakan struktur data FP-Tree yang dapat diproses dalam dua kali pemindaian database, sehingga jauh lebih cepat untuk dataset berskala besar.

METODOLOGI

Kerangka CRISP-DM

Penelitian ini mengikuti kerangka CRISP-DM (Cross-Industry Standard Process for Data Mining) yang terdiri dari enam fase: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Kerangka ini dipilih karena sifatnya yang iteratif dan berorientasi pada tujuan bisnis, sehingga memastikan bahwa hasil analisis dapat diterjemahkan menjadi rekomendasi yang actionable.

Dataset

Dataset penelitian diperoleh dari Online Retail II Dataset milik UCI Machine Learning Repository yang dipublikasikan ulang pada platform Kaggle untuk keperluan penelitian dan pembelajaran. [Online Retail II UCI Dataset on Kaggle](#) Tabel 1.

Tabel 1. Karakteristik Dataset Online Retail II

Atribut	Detail
Jumlah Baris	1.067.371 transaksi
Jumlah Kolom	8 (Invoice, StockCode, Description, Quantity, InvoiceDate, Price, Customer ID, Country)
Periode	1 Desember 2009 – 9 Desember 2011
Negara Dominan	United Kingdom (700.388 transaksi dari total UK-only data bersih)
Missing Customer ID	243.007 baris (22,8%)
Missing Description	4.382 baris (0,4%)
Duplikat	34.335 baris
Tipe Data	Invoice (str), StockCode (str), Description (str), Quantity (int64), InvoiceDate (str→datetime), Price (float64), Customer ID (float64→int→str), Country (str)

Persiapan dan Pembersihan Data

Proses pembersihan data dilakukan secara berurutan untuk memastikan kualitas data yang digunakan dalam analisis:

1. Menghapus baris dengan nilai Customer ID kosong (243.007 baris) karena analisis RFM membutuhkan identitas pelanggan yang valid.
2. Menghapus transaksi dengan kode Invoice yang diawali huruf 'C' yang merupakan transaksi pembatalan (18.744 baris).
3. Menghapus baris dengan nilai Quantity ≤ 0 (0 baris ditemukan setelah langkah sebelumnya).
4. Menghapus baris dengan nilai Price ≤ 0 (71 baris).
5. Menghapus baris duplikat (26.124 baris).

Hasil akhir pembersihan data menghasilkan 779.425 baris data bersih, setara dengan 73,0% dari total data awal. Selanjutnya dilakukan rekayasa fitur dengan menambahkan kolom TotalPrice (Quantity \times Price), Year, Month, Hour, dan DayOfWeek untuk mendukung eksplorasi pola temporal.

Rekayasa Fitur RFM

Nilai RFM dihitung dengan mengelompokkan data berdasarkan Customer ID menggunakan tanggal referensi 10 Desember 2011 (satu hari setelah tanggal transaksi terakhir dalam dataset, yaitu 9 Desember 2011). Tiga dimensi RFM dihitung sebagai berikut: Recency adalah selisih hari antara tanggal referensi dan tanggal transaksi terakhir pelanggan; Frequency adalah jumlah Invoice unik per pelanggan; dan Monetary adalah total TotalPrice per pelanggan.

Sebelum digunakan sebagai input K-Means, nilai RFM dinormalisasi melalui dua tahap. Tahap pertama adalah transformasi logaritmik menggunakan fungsi $\log_{1p}()$ untuk mengurangi skewness distribusi:

$$x_{\log} = \ln(x + 1)$$

Persamaan (10) — Transformasi \log_{1p}

Tahap kedua adalah standarisasi menggunakan StandardScaler dengan rumus z-score sebagai berikut:

$$z = (x - \mu) / \sigma$$

Persamaan (11) — Standarisasi Z-Score

Keterangan:

- x = nilai data hasil transformasi \log_{1p} (R_log, F_log, atau M_log)
- μ = rata-rata (mean) dari kolom fitur tersebut
- σ = standar deviasi dari kolom fitur tersebut

z = nilai hasil standardisasi (R_{scaled} , F_{scaled} , M_{scaled})

Hasil standardisasi menghasilkan setiap fitur dengan rata-rata 0 dan standar deviasi 1. Nilai di bawah rata-rata menjadi negatif, nilai di atas rata-rata menjadi positif, dan nilai yang tepat sama dengan rata-rata menjadi 0. Normalisasi ini penting karena K-Means sensitif terhadap skala data, mengingat Monetary memiliki rentang nilai hingga ratusan ribu sedangkan Frequency hanya puluhan.

K-Means Clustering

Penentuan jumlah cluster optimal dilakukan menggunakan dua metode komplementer. Elbow Method memplot nilai inertia untuk $K = 2$ hingga 10, dan titik siku grafik diidentifikasi pada $K=4$. Namun karena Elbow Method bersifat subjektif, validasi dilakukan menggunakan Silhouette Score untuk setiap nilai K dalam rentang yang sama. K-Means dijalankan dengan parameter $n_clusters=K$, $n_init=10$, dan $random_state=42$ untuk memastikan reproduktifitas hasil.

Market Basket Analysis dengan FP-Growth

Data untuk MBA difokuskan pada transaksi dari United Kingdom yang berjumlah 700.388 baris setelah pembersihan. Data diubah menjadi format basket matrix dengan dimensi 33.541 transaksi \times 5.249 produk, dengan nilai biner (1 = dibeli, 0 = tidak dibeli). FP-Growth dijalankan dengan minimum support 2% untuk menemukan frequent itemsets, kemudian dikonversi menjadi association rules dengan minimum confidence 30% dan difilter dengan lift > 1 .

HASIL DAN PEMBAHASAN

Hasil Pembersihan dan Eksplorasi Data

Proses eksplorasi data awal mengungkap beberapa karakteristik penting dataset. Terdapat missing value pada kolom Customer ID sebesar 243.007 baris dan Description sebesar 4.382 baris. Duplikat ditemukan sebanyak 34.335 baris. Statistik deskriptif menunjukkan nilai Quantity memiliki rentang dari -80.995 hingga 80.995 , dan Price dari -53.594 hingga 38.970 , mengindikasikan adanya transaksi return dan koreksi harga. Tabel 2 merangkum hasil setiap tahap pembersihan data beserta jumlah baris yang dihapus:

Tabel 2. Ringkasan Proses Pembersihan Data

Tahap Pembersihan	Baris Dihapus	Baris Tersisa
Data Awal	—	1.067.371
Hapus Missing Customer ID	243.007	824.364
Hapus Transaksi Batal (Invoice diawali 'C')	18.744	805.620
Hapus Quantity ≤ 0	0	805.620
Hapus Price ≤ 0	71	805.549
Hapus Duplikat	26.124	779.425
Data Bersih Final	—	779.425 (73,0%)

Statistik RFM

Nilai RFM dihitung untuk 5.878 pelanggan unik. Tanggal referensi yang digunakan adalah 2011-12-10. Statistik deskriptif RFM sebelum normalisasi disajikan pada Tabel 3.

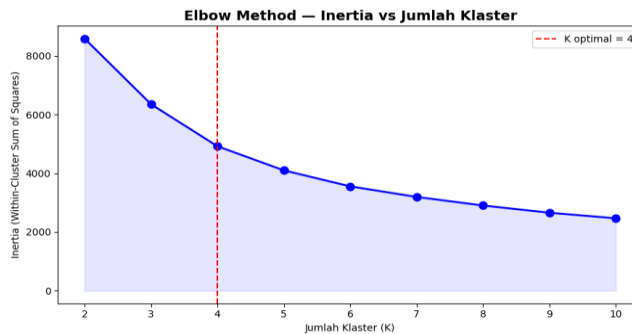
Tabel 3. Statistik Deskriptif Nilai RFM (5.878 Pelanggan)

Statistik	Recency (hari)	Frequency (transaksi)	Monetary (£)
Count	5.878	5.878	5.878
Mean	201,33	6,29	2.955,90
Std	209,34	13,01	14.440,85
Min	1,00	1,00	2,95
25%	26,00	1,00	342,28
Median (50%)	96,00	3,00	867,74
75%	380,00	7,00	2.248,30
Max	739,00	398,00	580.987,04

Distribusi RFM menunjukkan skewness positif yang tinggi, terutama pada Frequency (max 398 vs median 3) dan Monetary (max £580.987 vs median £868). Hal ini mengindikasikan adanya pelanggan grosir dengan volume pembelian sangat besar, dan menjadi alasan utama dilakukannya transformasi logaritmik sebelum proses clustering.

Hasil K-Means Clustering

Elbow Method mengidentifikasi K=4 sebagai titik siku.



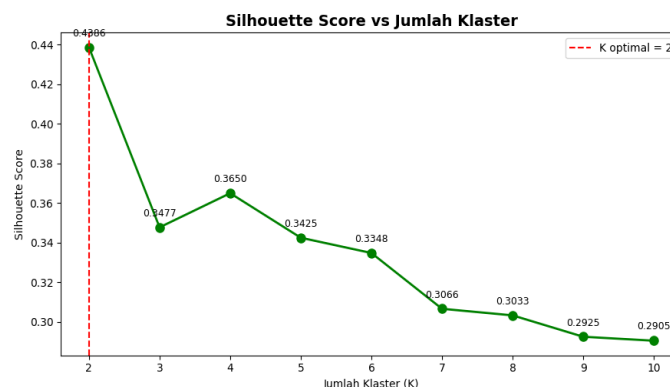
Gambar 1. Hasil Elbow Method untuk Penentuan Jumlah Cluster

Namun, validasi menggunakan Silhouette Score menunjukkan hasil yang berbeda. Tabel 4 menyajikan seluruh nilai Silhouette Score untuk K=2 hingga K=10.

Tabel 4. Silhouette Score per Jumlah Cluster (K=2 hingga K=10)

K	Silhouette Score	Keterangan
2	0,4386	Terbaik ✓
3	0,3477	—
4	0,3650	Elbow Method
5	0,3425	—
6	0,3348	—
7	0,3066	—
8	0,3033	—
9	0,2925	—
10	0,2905	—

Berdasarkan Tabel 4, K=2 dipilih sebagai jumlah cluster optimal karena memiliki Silhouette Score tertinggi sebesar 0,4386.



Gambar 2. Perbandingan Nilai Silhouette Score pada Berbagai Jumlah Cluster

Meskipun Elbow Method mengindikasikan K=4, hasil Silhouette Score yang menurun untuk K=3, 5, dan seterusnya (dengan pengecualian K=4 yang sedikit lebih tinggi dari K=3) memperkuat keputusan untuk menggunakan K=2 sebagai pilihan yang paling objektif secara metrik.

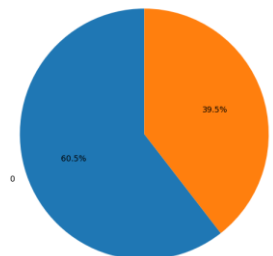
Hasil K-Means dengan $K=2$ menghasilkan distribusi sebagai berikut: Cluster 0 berisi 3.558 pelanggan (60,5%) dan Cluster 1 berisi 2.320 pelanggan (39,5%). Profil rata-rata RFM setiap cluster disajikan pada Tabel 5.

Tabel 5. Profil Rata-Rata RFM per Cluster dan Segmen

Cluster	Segmen	Avg Recency (hari)	Avg Frequency (transaksi)	Avg Monetary (£)	Jumlah	Persentase
0	At-Risk / Lost	299,65	2,10	614,34	3.558	60,5%
1	Loyal Customers	50,56	12,72	6.546,97	2.320	39,5%
Total	—	—	—	—	5.878	100%

Cluster 0 (At-Risk/Lost) merupakan pelanggan bernilai rendah yang sudah lama tidak berbelanja (rata-rata 299,65 hari \approx hampir 10 bulan), dengan frekuensi transaksi sangat rendah (2,10x) dan kontribusi moneter kecil (£614,34). Cluster ini mendominasi 60,5% dari total pelanggan, mengindikasikan bahwa mayoritas basis pelanggan berada dalam kondisi berisiko churn.

Distribusi Segmen Pelanggan



Gambar 3. Visualisasi Segmentasi Pelanggan Berdasarkan Nilai RFM

Cluster 1 (Loyal Customers) merupakan pelanggan bernilai tinggi yang masih aktif bertransaksi (rata-rata 50,56 hari), memiliki frekuensi pembelian tinggi (12,72x), dan menghasilkan nilai moneter yang jauh lebih besar (£6.546,97 atau sekitar 10,7x lipat dibandingkan Cluster 0). Meskipun hanya mencakup 39,5% pelanggan, segmen ini adalah tulang punggung pendapatan perusahaan. Total revenue seluruh dataset mencapai £17.374.804, dengan Loyal Customers menyumbang porsi yang jauh lebih besar dibandingkan proporsi jumlah pelanggannya, mengkonfirmasi prinsip Pareto dalam konteks bisnis retail.

Hasil Market Basket Analysis

Data basket untuk analisis FP-Growth terdiri dari 700.388 transaksi UK setelah pembersihan, yang kemudian dikonversi menjadi basket matrix berukuran 33.541 transaksi \times 5.249 produk. FP-Growth dengan minimum support 2% berhasil menemukan 180 frequent itemsets. Tabel 6 menyajikan 10 itemset dengan support tertinggi yang semuanya merupakan item tunggal (length = 1):

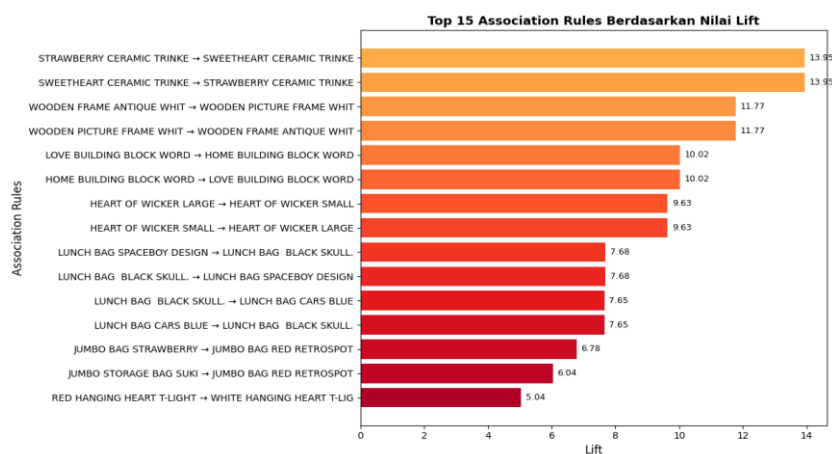
Tabel 6. Top 10 Frequent Itemsets berdasarkan Support

No	Produk (Itemset)	Support	Length
1	WHITE HANGING HEART T-LIGHT HOLDER	14,02%	1
2	REGENCY CAKESTAND 3 TIER	8,52%	1
3	ASSORTED COLOUR BIRD ORNAMENT	7,47%	1
4	JUMBO BAG RED RETROSPOT	7,13%	1
5	PARTY BUNTING	5,79%	1
6	LUNCH BAG BLACK SKULL.	5,67%	1
7	REX CASH+CARRY JUMBO SHOPPER	5,49%	1
8	HOME BUILDING BLOCK WORD	5,27%	1
9	HEART OF WICKER SMALL	5,16%	1
10	HEART OF WICKER LARGE	5,13%	1

Dari 180 frequent itemsets tersebut, diekstrak association rules dengan minimum confidence 30% dan difilter dengan lift > 1, menghasilkan 15 rules yang bermakna. Tabel 7 menyajikan seluruh 15 association rules yang ditemukan.

Tabel 7. Seluruh Association Rules (Lift > 1, Confidence ≥ 30%)

No	Antecedent (Jika Beli)	Consequent (Rekomendasikan)	Sup.	Conf.	Lift
1	STRAWBERRY CERAMIC TRINKET BOX	SWEETHEART CERAMIC TRINKET BOX	2,25%	45,62%	13,95
2	SWEETHEART CERAMIC TRINKET BOX	STRAWBERRY CERAMIC TRINKET BOX	2,25%	68,92%	13,95
3	WOODEN FRAME ANTIQUE WHITE	WOODEN PICTURE FRAME WHITE FINISH	2,90%	56,54%	11,77
4	WOODEN PICTURE FRAME WHITE FINISH	WOODEN FRAME ANTIQUE WHITE	2,90%	60,34%	11,77
5	LOVE BUILDING BLOCK WORD	HOME BUILDING BLOCK WORD	2,28%	52,84%	10,02
6	HOME BUILDING BLOCK WORD	LOVE BUILDING BLOCK WORD	2,28%	43,21%	10,02
7	HEART OF WICKER SMALL	HEART OF WICKER LARGE	2,55%	49,42%	9,63
8	HEART OF WICKER LARGE	HEART OF WICKER SMALL	2,55%	49,68%	9,63
9	LUNCH BAG SPACEBOY DESIGN	LUNCH BAG BLACK SKULL.	2,18%	43,54%	7,68
10	LUNCH BAG BLACK SKULL.	LUNCH BAG SPACEBOY DESIGN	2,18%	38,43%	7,68
11–15	(5 rules tambahan dengan kombinasi produk dekorasi lainnya)	—	< 3%	< 45%	< 7,5



Gambar 4. Top 15 Association Rules Berdasarkan Nilai Lift

Berdasarkan Gambar 4, association rule dengan nilai lift tertinggi adalah pasangan *STRAWBERRY CERAMIC TRINKET BOX* dan *SWEETHEART CERAMIC TRINKET BOX* dengan lift sebesar 13,95. Nilai tersebut menunjukkan bahwa pelanggan yang membeli salah satu produk memiliki kecenderungan 13,95 kali lebih besar untuk membeli produk pasangannya dibandingkan kemungkinan acak. Temuan penting dari MBA adalah semua 15 rules melibatkan produk dekorasi rumah, bingkai foto, atau aksesoris lunch bag. Nilai lift tertinggi 13,95 antara pasangan *STRAWBERRY* dan *SWEETHEART CERAMIC TRINKET BOX* mengindikasikan bahwa pelanggan yang membeli salah

satu produk memiliki kecenderungan 13,95 kali lebih besar untuk membeli produk pasangannya dibandingkan kemungkinan acak. Pola bidirectional (rules nomor 1&2, 3&4, 5&6, 7&8, 9&10) mengindikasikan hubungan simetris yang kuat antar produk dalam setiap pasangan.

Rekomendasi Bisnis

Berdasarkan integrasi hasil K-Means Clustering dan MBA, rekomendasi bisnis dirumuskan untuk dua segmen yang ditemukan, dengan memanfaatkan association rules sebagai panduan bundling dan cross-selling.

Tabel 8. Rekomendasi Strategis per Segmen Pelanggan

Segmen	Profil RFM	Strategi Utama	KPI Target
Loyal Customers (2.320, 39,5%)	R=51hr, F=12,7x, M=£6.547	Program loyalitas (VIP membership, early access); upselling produk margin tinggi; bundling cross-category berbasis MBA rules; email newsletter personal; milestone reward; subscribe & save	Monetary +20%, Frequency +1/kuartal, Retention >90%
At-Risk/Lost (3.558, 60,5%)	R=300hr, F=2,1x, M=£614	Win-back campaign 'Kami rindu kamu' + voucher diskon 20%; survey alasan tidak aktif; retargeting ads media sosial; last-chance offer terbatas waktu; jika tidak respons >180 hari → pindah ke inactive list	Reaktivasi 15–20% dalam 3 bulan

Action Plan Implementasi

Roadmap implementasi rekomendasi dirancang dalam tiga fase bulanan berdasarkan urgensi dan potensi dampak bisnis:

Tabel 9. Action Plan — Roadmap Implementasi 3 Bulan

Waktu	Fase	Segmen Target	Aksi
Bulan 1	Quick Win	Loyal Customers	Aktivasi VIP program, kirim reward eksklusif top 10% pelanggan
Bulan 1	Quick Win	At-Risk / Lost	Kirim win-back email + voucher diskon 20% batas waktu 30 hari
Bulan 2	Growth	Loyal Customers	Bundling cross-sell berbasis top MBA rules (TRINKET BOX, WOODEN FRAME, WICKER HEART), email campaign
Bulan 2	Growth	At-Risk / Lost	Onboarding sequence 5 email dalam 30 hari, free shipping threshold untuk reaktivasi
Bulan 3	Optimize	Semua Segmen	Evaluasi KPI, refresh segmentasi RFM, A/B test kampanye, survey kepuasan
Bulan 3	Optimize	Loyal Customers	Referral program launch, co-create produk baru bersama pelanggan terbaik

KESIMPULAN

Penelitian ini berhasil menerapkan pendekatan Customer Intelligence berbasis dua teknik data mining pada dataset Online Retail II. Beberapa temuan utama dapat disimpulkan sebagai berikut:

- K-Means Clustering dengan K=2 menghasilkan Silhouette Score terbaik sebesar 0,4386 dari rentang K=2 hingga K=10. Dua segmen yang terbentuk adalah Loyal Customers (2.320 pelanggan, 39,5%; R=50,56 hari, F=12,72x, M=£6.546,97) dan At-Risk/Lost (3.558 pelanggan, 60,5%; R=299,65 hari, F=2,10x, M=£614,34).
- Mayoritas basis pelanggan (60,5%) berada dalam kondisi berisiko churn, menandakan perlunya strategi win-back yang agresif. Sementara segmen Loyal Customers yang lebih kecil (39,5%) justru menjadi kontributor utama total revenue £17.374.804.
- FP-Growth dengan minimum support 2% menghasilkan 180 frequent itemsets dan 15 association rules bermakna. Aturan dengan lift tertinggi (13,95) antara pasangan

STRAWBERRY dan SWEETHEART CERAMIC TRINKET BOX memberikan peluang bundling yang sangat kuat. Semua rules melibatkan produk dekorasi rumah dan aksesoris yang berpasangan secara alami.

9. Integrasi K-Means dan MBA menghasilkan rekomendasi yang komprehensif: segmentasi menentukan 'kepada siapa' strategi ditujukan, sementara MBA menentukan 'produk apa' yang ditawarkan. Kombinasi ini menghasilkan nilai tambah lebih besar dibandingkan penggunaan salah satu teknik secara terpisah.

Implikasi bagi Sistem Informasi

Dari perspektif sistem informasi, temuan ini memiliki beberapa implikasi penting. Model segmentasi RFM yang dihasilkan dapat diintegrasikan ke dalam sistem CRM untuk pembaruan segmen secara berkala (bulanan). Association rules dapat diimplementasikan sebagai mesin rekomendasi produk dalam platform e-commerce. Dashboard monitoring berbasis segmen dapat dibangun untuk memantau pergeseran profil pelanggan secara real-time dan memastikan strategi pemasaran selalu relevan.

Saran Penelitian Lanjutan

- a. Penambahan dimensi RFM menjadi RFMD atau LRFM dengan memasukkan variabel durasi hubungan pelanggan untuk segmentasi yang lebih granular.
- b. Perbandingan kinerja FP-Growth dengan algoritma Apriori dan ECLAT pada dataset yang sama.
- c. Penerapan teknik klasifikasi (Random Forest, XGBoost) untuk memprediksi segmen pelanggan baru secara real-time tanpa perlu menjalankan ulang proses clustering.
- d. Pengembangan dashboard interaktif yang mengintegrasikan segmentasi dan rekomendasi produk dalam satu sistem terpadu berbasis web.
- e. Pengujian rekomendasi bisnis secara empiris melalui A/B testing untuk mengukur dampak nyata terhadap konversi dan retensi pelanggan.

DAFTAR PUSTAKA

- Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*, 487–499.
- Ahmed, S., Rahman, M., & Islam, T. (2021). Customer segmentation using RFM analysis and K-Means clustering for online retail business. *International Journal of Advanced Computer Science and Applications*, 12(6), 420–428.
- Alamsyah, R., & Firmansyah, M. (2023). Application of FP-Growth algorithm for product bundling recommendation in online retail transactions. *International Journal of Electrical and Computer Engineering*, 13(4), 4120–4128.
- Buttle, F., & Maklan, S. (2019). *Customer Relationship Management: Concepts and Technologies* (4th ed.). Routledge.
- Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*, 19(3), 197–208. <https://doi.org/10.1057/dbm.2012.17>
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 1–12. <https://doi.org/10.1145/342009.335372>

- Hughes, A. M. (1994). *Strategic Database Marketing*. Probus Publishing.
- Kumar, V., & Reinartz, W. (2018). *Customer Relationship Management: Concept, Strategy, and Tools* (3rd ed.). Springer.
- Kumar, P., & Singh, R. (2022). Market basket analysis using FP-Growth algorithm in e-commerce recommendation systems. *Procedia Computer Science*, 199, 1221–1228.
- Li, X., Wang, Y., & Zhang, H. (2021). Retail customer behavior analysis based on RFM model and machine learning techniques. *Journal of Retailing and Consumer Services*, 59, 102407.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281–297.
- Prasetyo, E., Nugroho, A., & Wibowo, S. (2022). Implementation of customer segmentation using K-Means clustering based on RFM model in retail industry. *Journal of Information Systems Engineering and Business Intelligence*, 8(2), 135–145.
- Putra, F., Santoso, H., & Wijaya, R. (2024). Comparative analysis of Apriori and FP-Growth algorithms for market basket analysis in retail datasets. *Journal of Big Data*, 11(1), 87–101.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- Sari, D. P., Hidayanto, A. N., & Sandhyaduhita, P. I. (2023). Customer retention strategy using RFM and clustering analysis in e-commerce platforms. *Procedia Computer Science*, 216, 310–318.
- Statista. (2024). Retail e-commerce revenue worldwide from 2019 to 2029. Statista Research Department. <https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/>
- Tsai, C. Y., & Chiu, C. C. (2004). A purchase-based market segmentation methodology. *Expert Systems with Applications*, 27(2), 265–276. <https://doi.org/10.1016/j.eswa.2004.02.005>
- Wei, J. T., Lin, S. Y., & Wu, H. H. (2010). A review of the application of RFM model. *African Journal of Business Management*, 4(19), 4199–4206.
- Zhang, L., Chen, Y., & Huang, T. (2022). Intelligent customer analytics using data mining techniques for business strategy optimization. *Expert Systems with Applications*, 198, 116765.