



Segmentasi Pelanggan dan Analisis Pola Pembelian Berbasis Integrasi K-Means dan FP-Growth pada Data Transaksi Ritel

Aditya Saputra¹, Dadang Priyanto², Galih Hendro Martono³

¹Program Studi Teknologi Informasi, Fakultas Teknik, Universitas Bumigora, Indonesia

^{2,3}Departement Of Computer Science, Master Program, Universitas Bumigora

aditputra02@gmail.com¹, dadang.priyanto@universitasbumigora.ac.id²

galih.hendro@universitasbumigora.ac.id³

Korespondensi penulis: aditputra02@gmail.com

Abstract. *Advances in information technology in the retail sector have generated a massive volume of transaction data, which can be leveraged to evaluate consumer behavior and support business decision-making. By combining the K-Means and FP-Growth algorithms on a retail transaction dataset, this study aims to develop a product recommendation system based on purchasing patterns, using a dataset consisting of 1,000,000 transaction records from the 2020–2024 period obtained from Kaggle. The research methodology used is CRISP-DM. Customers were grouped based on transaction behavior characteristics using the K-Means algorithm subsequently, association rules and purchasing patterns were analyzed using the FP-Growth algorithm. Based on the elbow method, three customer clusters were identified: Cluster 0 consists of customers with low shopping activity Cluster 1 consists of active and loyal customers and Cluster 2 consists of customers who are sensitive to promotions and discounts. The evaluation results show that of the three scenarios tested, Scenario II generated 160 association rules for each cluster while still producing frequent item sets, resulting in a more manageable number of rules that can still be used for purchasing pattern analysis. The findings of this study indicate that the combination of K-Means and FP-Growth can generate recommendations based on customer characteristics and support marketing and sales strategies.*

Keywords: *Data Mining, K-Means, FP-Growth, Sistem Rekomendasi, Retail, Flask*

Abstrak Kemajuan teknologi informasi di sektor ritel menghasilkan data transaksi dalam jumlah besar yang dapat dimanfaatkan untuk memahami perilaku pelanggan dan mendukung pengambilan keputusan bisnis berbasis data. Penelitian ini bertujuan menganalisis pola pembelian dan mengembangkan sistem rekomendasi produk melalui integrasi algoritma *K-Means* dan *FP-Growth*. Penelitian dilaksanakan menggunakan tahapan CRISP-DM dengan melibatkan 1.000.000 catatan transaksi ritel periode 2020–2024 yang diperoleh dari Kaggle. *K-Means* diterapkan untuk mengelompokkan pelanggan berdasarkan jumlah transaksi, total pengeluaran, rata-rata pengeluaran, jumlah produk yang dibeli, serta kecenderungan penggunaan promosi dan diskon. Selanjutnya, *FP-Growth* digunakan untuk menemukan *frequent itemsets* dan membentuk aturan asosiasi pada setiap kluster. Hasil metode *elbow* menunjukkan bahwa jumlah kluster yang sesuai adalah tiga. Kluster 0 terdiri atas 188.607 pelanggan dengan aktivitas belanja relatif rendah, Kluster 1 mencakup 3.171 pelanggan aktif dan loyal, sedangkan Kluster 2 terdiri atas 137.960 pelanggan yang cenderung responsif terhadap promosi dan diskon. Penerapan *FP-Growth* dengan minimum *support* sebesar 0,001 dan minimum *confidence* sebesar 0,01 menghasilkan 19.222 aturan asosiasi, yaitu 6.478 aturan pada Kluster 0, 6.312 aturan pada Kluster 1, dan 6.432 aturan pada Kluster 2. Sebagian besar aturan memiliki nilai *lift* di bawah 1, yang menunjukkan bahwa keterkaitan antarproduk masih relatif lemah. Meskipun demikian, pada Kluster 1 ditemukan aturan *Onions* → *Toothpaste* dengan nilai *lift* sebesar 1,014691, yang mengindikasikan hubungan positif antara kedua produk. Temuan ini menunjukkan bahwa integrasi *K-Means* dan *FP-Growth* dapat mengidentifikasi pola pembelian berdasarkan karakteristik setiap segmen pelanggan. Hasil analisis tersebut telah diimplementasikan dalam aplikasi berbasis web dan dapat digunakan sebagai informasi pendukung dalam merancang strategi pemasaran dan penjualan yang lebih terarah.

Keywords: segmentasi pelanggan, K-Means, FP-Growth, pola pembelian, sistem rekomendasi

1. LATAR BELAKANG

Pendahuluan

Perkembangan teknologi dan perubahan perilaku konsumen telah mendorong industri ritel untuk beradaptasi dengan lingkungan bisnis yang semakin dinamis. Persaingan tidak lagi hanya ditentukan oleh ketersediaan produk dan harga, tetapi juga oleh kemampuan perusahaan dalam memahami kebutuhan pelanggan serta merespons perubahan pasar secara cepat. Kondisi tersebut menuntut pelaku usaha ritel untuk meningkatkan efisiensi operasional sekaligus mempertahankan daya saing melalui pemanfaatan teknologi informasi (Nikmah et al., 2025). Salah satu pendekatan yang dapat digunakan adalah *data mining*, yaitu proses pengolahan data dalam jumlah besar untuk menemukan pola dan informasi yang dapat mendukung pengambilan keputusan bisnis (Bunda, 2020).

Di sisi lain, industri ritel juga menghadapi sejumlah persoalan, mulai dari tingginya tingkat persaingan, perubahan pola konsumsi masyarakat, hingga pengelolaan persediaan dan rantai pasok (Rifda, 2023). Pada sisi konsumen, kemudahan memperoleh informasi memberikan lebih banyak pilihan dalam menentukan tempat berbelanja. Konsumen dapat mempertimbangkan harga, variasi dan ketersediaan produk, lokasi, promosi, maupun bentuk pelayanan yang diberikan sebelum melakukan pembelian. Situasi tersebut mendorong perusahaan ritel untuk merancang strategi pemasaran yang tidak hanya berorientasi pada peningkatan omzet, tetapi juga mampu menyesuaikan penawaran dengan karakteristik dan kebutuhan konsumennya (Jadid, 2021).

Salah satu sumber informasi yang dapat dimanfaatkan untuk memahami perilaku tersebut adalah data transaksi penjualan. Setiap transaksi menyimpan informasi mengenai produk yang dibeli, frekuensi transaksi, nilai pembelian, penggunaan diskon, serta keterlibatan pelanggan dalam program promosi. Apabila dianalisis secara sistematis, data tersebut dapat memberikan gambaran mengenai kecenderungan dan keterkaitan produk yang muncul dalam aktivitas belanja konsumen. Analisis pola pembelian menjadi penting karena dapat membantu perusahaan menemukan kombinasi produk yang sering dibeli secara bersamaan dan memanfaatkannya sebagai dasar dalam penyusunan strategi pemasaran maupun rekomendasi produk (Hairani & Ximenes Guterres, 2024). Pemanfaatan data yang tepat juga memungkinkan pengambilan keputusan dilakukan

berdasarkan informasi yang lebih terukur sehingga dapat mendukung peningkatan penjualan dan daya saing perusahaan (Arisandi et al., 2023).

Sejumlah penelitian telah memanfaatkan teknik asosiasi untuk menggali pola pembelian konsumen. Valencia & Tisno Atmojo (2024) mengidentifikasi produk yang sering dibeli serta memahami perilaku konsumen berdasarkan data transaksi bulanan yang diolah menggunakan RapidMiner. Mardiana, N. A., & Windari (2024) menggunakan algoritma *Frequent Pattern Growth* (FP-Growth) untuk menganalisis pola penjualan pada Lathansa Cafe & Ramen. Pendekatan serupa digunakan oleh Saraswati et al (2024) pada data ritel non-toko dengan menerapkan kerangka kerja CRISP-DM dan FP-Growth pada tahap pemodelan. Sementara itu, Suhada et al (2020) menerapkan FP-Growth pada data transaksi AHASS Cibadak untuk menemukan pola pembelian konsumen.

Hasil berbagai penelitian tersebut menunjukkan bahwa FP-Growth dapat digunakan untuk menemukan *frequent itemset* dan hubungan antar produk berdasarkan kecenderungan produk muncul secara bersamaan dalam transaksi. Meskipun demikian, pendekatan yang hanya menganalisis keseluruhan transaksi sebagai satu kelompok memiliki keterbatasan karena pola pembelian setiap pelanggan belum tentu bersifat homogen. Pelanggan dengan frekuensi transaksi rendah, pelanggan dengan intensitas pembelian tinggi, maupun pelanggan yang sering memanfaatkan promosi dapat memiliki kecenderungan pembelian yang berbeda. Apabila seluruh transaksi dianalisis secara global tanpa mempertimbangkan karakteristik pelanggan, pola asosiasi yang dihasilkan berpotensi lebih merepresentasikan kecenderungan umum dan belum menggambarkan pola pembelian yang spesifik pada kelompok pelanggan tertentu.

Berdasarkan kondisi tersebut, penelitian ini mengombinasikan algoritma K-Means dan FP-Growth untuk menghasilkan analisis pola pembelian berbasis segmentasi pelanggan. K-Means digunakan terlebih dahulu untuk mengelompokkan pelanggan berdasarkan kemiripan karakteristik perilaku transaksi. Karakteristik tersebut direpresentasikan melalui sejumlah atribut yang diperoleh dari proses agregasi data transaksi, seperti frekuensi transaksi, total dan rata-rata pengeluaran, jumlah produk yang dibeli, serta keterlibatan pelanggan terhadap diskon dan program promosi. Pendekatan ini memungkinkan pelanggan dengan karakteristik transaksi yang relatif serupa ditempatkan pada kelompok yang sama.

Setelah proses segmentasi dilakukan, FP-Growth diterapkan secara terpisah pada transaksi di masing-masing *cluster*. Pendekatan tersebut memungkinkan proses pencarian *frequent itemset* dan *association rules* dilakukan berdasarkan karakteristik kelompok pelanggan, bukan hanya berdasarkan pola transaksi secara keseluruhan. Dengan demikian, aturan asosiasi yang diperoleh dapat digunakan untuk mengidentifikasi kecenderungan kombinasi produk yang lebih spesifik pada setiap segmen pelanggan.

2. KAJIAN TEORITIS

Algoritma K-Means

K-Means adalah teknik pengelompokan data di mana keanggotaan setiap titik data dalam suatu kelompok ditentukan oleh tingkat keanggotaan. Data tersebut kemudian akan dikelompokkan berdasarkan kesamaan tingkat keanggotaan data di dalam kumpulan data (Makhfuzi Fahmi, Zumrotun Naimah, Novi Setiani, 2014). Adapun fungsi algoritma *K-Means Clustering* di antaranya segmentasi data, pengurangan dimensi, pengelompokan data dan pengenalan pola (Annisa, 2026). Selain itu prosedur perhitungan algoritma *K-Means* di antaranya, menentukan jumlah *cluster* (k), menentukan titik pusat *cluster* (*centroid*), dan menghitung jarak data terhadap *centroid* (Rohmah et al., 2021).

Algoritma FP-Growth

Algoritma *FP-Growth* merupakan pengembangan dari algoritma Apriori, sehingga kelemahan algoritma Apriori diperbaiki oleh algoritma *FP-Growth*. Dalam algoritma Apriori, proses pembangkitan kandidat diperlukan untuk mendapatkan himpunan item yang sering, namun, dalam algoritma *FP-Growth*, proses pembangkitan kandidat tidak dilakukan karena telah menggunakan konsep konstruksi pohon dalam analisis *FP-tree* untuk himpunan item yang sering. Dengan menggunakan *FP-tree*, himpunan item yang sering dapat diekstraksi dengan mudah dari *FP-tree*. (Suhada et al. 2020).

Adapun rumus support dan confidence sebagai berikut : (M. Rizky, A. Utama, 2020)

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

$$\text{Confidence (A} \rightarrow \text{B)} = \frac{P(\text{A}|\text{B})}{\frac{\Sigma \text{Transaksi mengandung A dan B}}{\Sigma \text{Transaksi mengandung A}}}$$

Data Mining

Data Mining merupakan suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari sebuah data (Journal et al., 2020) dan serangkaian prosedur yang dirancang untuk mengungkap informasi tambahan yang saat ini tidak ada dalam database, Prosesnya memerlukan pola eksplorasi dan model data yang mampu mengubah data mentah menjadi wawasan yang berharga.(H. Lestari et al., 2024). Fungsi dan peran data mining dalam perusahaan terdiri dari optimalisasi pengambilan keputusan, peningkatan pengalaman pelanggan, inovasi produk dan layanan, efisiensi operasional Setiyani et al. (2025).

3. METODE PENELITIAN

Penelitian ini menggunakan kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM) sebagai acuan dalam mengolah data transaksi hingga memperoleh pola pembelian pelanggan. CRISP-DM dipilih karena menyediakan tahapan analisis data yang sistematis dan memungkinkan setiap proses penelitian dilakukan secara berurutan, mulai dari pemahaman terhadap permasalahan, pengenalan karakteristik data, persiapan data, pemodelan, evaluasi hasil, hingga implementasi model (Hidayatullah et al., 2025).

Secara umum, penelitian dilaksanakan melalui enam tahapan CRISP-DM, yaitu *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment*. K-Means digunakan pada tahap pemodelan untuk membentuk segmentasi pelanggan berdasarkan karakteristik transaksi, sedangkan FP-Growth diterapkan pada setiap *cluster* untuk menemukan *frequent itemset* dan aturan asosiasi antar produk.

3.1 Business Understanding

Penelitian ini berangkat dari kebutuhan untuk memperoleh dua jenis informasi dari data transaksi. Pertama, mengidentifikasi kelompok pelanggan yang memiliki kemiripan perilaku transaksi. Kedua, menemukan kecenderungan produk yang dibeli secara bersamaan pada setiap kelompok pelanggan. Informasi tersebut dapat digunakan

untuk memberikan gambaran mengenai karakteristik pelanggan dan pola pembelian yang terjadi pada masing-masing segmen.

3.2 Data Understanding

Data penelitian diperoleh dari *Retail Transactions Dataset* yang tersedia melalui platform Kaggle. Dataset terdiri atas 1.000.000 catatan transaksi dalam periode 2020–2024. Setiap baris pada data awal merepresentasikan satu aktivitas transaksi sehingga pelanggan yang melakukan transaksi lebih dari satu kali dapat muncul pada beberapa baris berbeda.

Atribut yang digunakan meliputi identitas transaksi (*Transaction_ID*), waktu transaksi (*Date*), identitas pelanggan (*Customer_Name*), daftar produk yang dibeli (*Product*), penggunaan diskon (*Discount_Applied*), dan jenis promosi (*Promotion*). Atribut-atribut tersebut memberikan informasi mengenai siapa yang melakukan transaksi, produk yang dibeli, serta keterlibatan pelanggan terhadap diskon dan program promosi.

3.3 Data Preparation

Tahap *Data Preparation* dilakukan untuk mengubah data transaksi awal menjadi representasi yang sesuai dengan kebutuhan pemodelan. Persiapan data meliputi proses pengkodean atribut kategorikal dan agregasi transaksi berdasarkan pelanggan.

3.3.1 Encoding Data Kategorikal

Beberapa atribut pada dataset memiliki nilai kategorikal sehingga perlu dikonversi ke dalam bentuk numerik sebelum digunakan dalam proses clustering. Pengkodean diterapkan pada atribut *Discount_Applied* dan *Promotion*.

Pada atribut *Discount_Applied*, nilai *True* dikonversi menjadi 1 untuk menunjukkan transaksi yang menggunakan diskon, sedangkan nilai *False* dikonversi menjadi 0 untuk transaksi tanpa diskon. Sementara itu, atribut *Promotion* dikonversi berdasarkan kategori promosi yang terdapat pada dataset, yaitu *None* = 0, *BOGO* (Buy One Get One) = 1, dan *Discount On Selected Items* = 2. Hasil pengkodean menghasilkan representasi numerik yang selanjutnya digunakan pada proses agregasi data pelanggan.

3.3.2 Agregasi Data Pelanggan

Dataset awal terdiri atas 1.000.000 catatan transaksi yang masih merepresentasikan aktivitas pembelian secara individual. Kondisi tersebut menyebabkan satu pelanggan dapat muncul pada beberapa baris apabila melakukan transaksi lebih dari satu kali. Oleh karena itu, dilakukan agregasi berdasarkan identitas pelanggan agar data

yang digunakan dalam proses clustering dapat menggambarkan karakteristik transaksi setiap pelanggan secara lebih terstruktur.

3.4 Modeling

Tahap *Modeling* terdiri atas dua proses utama, yaitu segmentasi pelanggan menggunakan K-Means dan pencarian pola pembelian menggunakan FP-Growth. Kedua algoritma diterapkan secara berurutan sehingga hasil clustering menjadi dasar pengelompokan data sebelum proses pencarian aturan asosiasi dilakukan.

3.4.1 Segmentasi Pelanggan Menggunakan K-Means

K-Means digunakan untuk mengelompokkan pelanggan berdasarkan kemiripan karakteristik transaksi yang diperoleh dari tahap agregasi. Proses *clustering* dilakukan menggunakan atribut *transaction_count*, *total_spent*, *avg_spent*, *total_items*, *avg_items*, *discount_ratio*, dan *promo_ratio*. Secara umum, *K-Means* bekerja dengan menentukan sejumlah pusat cluster (*centroid*) dan menempatkan setiap data ke *cluster* dengan *centroid* terdekat. Kedekatan data terhadap *centroid* ditentukan berdasarkan jarak *Euclidean*. Penentuan jumlah *cluster* dilakukan menggunakan *Elbow Method* dengan mengamati perubahan nilai *Within-Cluster Sum of Squares* (WCSS)

Jumlah *cluster* dipilih pada titik ketika penurunan WCSS mulai melandai dan tidak lagi menunjukkan perubahan yang signifikan. Berdasarkan proses tersebut, jumlah cluster yang digunakan dalam penelitian ditetapkan untuk proses segmentasi pelanggan.

3.4.2 Analisis Pola Pembelian Menggunakan FP-Growth

Setelah setiap pelanggan memperoleh label cluster, transaksi dikelompokkan berdasarkan keanggotaan *cluster* tersebut. *FP-Growth* selanjutnya diterapkan secara terpisah pada masing-masing *cluster* untuk mengidentifikasi kombinasi produk yang sering muncul secara bersamaan.

Pengujian *FP-Growth* dilakukan menggunakan beberapa kombinasi nilai minimum *support* dan minimum *confidence*. Skenario tersebut digunakan untuk mengamati perubahan jumlah *frequent itemset* dan *association rules* yang dihasilkan pada setiap cluster. Skenario pengujian FP-Growth dapat dilihat pada Tabel 1.

Tabel 1. Skenario Pengujian FP-Growth

Skenario	Cluster	Minimum Support	Minimum Confidence
I	0, 1, 2	0,1%	0,1%
II	0, 1, 2	0,2%	0,2%

III	0, 1, 2	30%	50%
-----	---------	-----	-----

3.5 Evaluation

Tahap *Evaluation* dilakukan untuk menilai hasil segmentasi pelanggan dan pola asosiasi yang diperoleh dari proses pemodelan. Pada *K-Means*, evaluasi awal dilakukan melalui *Elbow Method* dengan membandingkan nilai WCSS pada beberapa alternatif jumlah *cluster*. Titik siku pada grafik digunakan sebagai dasar dalam menentukan jumlah *cluster* yang digunakan pada tahap clustering.

Evaluasi *FP-Growth* dilakukan dengan membandingkan hasil dari ketiga skenario minimum *support* dan minimum *confidence*. Parameter yang diamati meliputi jumlah transaksi pada setiap *cluster*, jumlah *frequent itemset* yang ditemukan, serta jumlah *association rules* yang dihasilkan. Perbandingan tersebut digunakan untuk melihat pengaruh perubahan nilai ambang terhadap banyaknya pola asosiasi yang dapat ditemukan dari transaksi pada masing-masing kelompok pelanggan.

3.6 Deployment

Tahap terakhir CRISP-DM adalah *Deployment*. Pada tahap ini, proses pengolahan data, segmentasi pelanggan, dan analisis pola pembelian diintegrasikan ke dalam sistem berbasis web menggunakan framework Flask. Implementasi berbasis web digunakan sebagai bentuk penerapan hasil analisis agar proses segmentasi dan eksplorasi pola pembelian tidak hanya berhenti pada tahap eksperimen, tetapi dapat diakses melalui antarmuka aplikasi yang lebih mudah digunakan.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Persiapan dan Agregasi Data

Dataset awal yang digunakan dalam penelitian terdiri atas 1.000.000 catatan transaksi yang merepresentasikan aktivitas pembelian pelanggan selama periode pengamatan. Pada data awal, satu pelanggan dapat tercatat pada beberapa transaksi yang berbeda. Kondisi tersebut menyebabkan data belum dapat digunakan secara langsung untuk menggambarkan karakteristik perilaku masing-masing pelanggan pada proses *clustering*. Oleh karena itu, transaksi terlebih dahulu diagregasikan berdasarkan identitas pelanggan. Proses agregasi menghasilkan satu representasi untuk setiap pelanggan dengan sejumlah atribut yang menggambarkan aktivitas transaksinya, meliputi *transaction_count*, *total_spent*, *avg_spent*, *total_items*, *avg_items*, *discount_ratio*, dan

promo_ratio. Setelah proses tersebut dilakukan, 1.000.000 catatan transaksi menghasilkan 329.738 data pelanggan yang selanjutnya digunakan sebagai unit analisis pada proses K-Means. Contoh dari dataset yang digunakan dapat dilihat pada Tabel 2.

Tabel 2. Sampel Hasil Agregasi Data Pelanggan

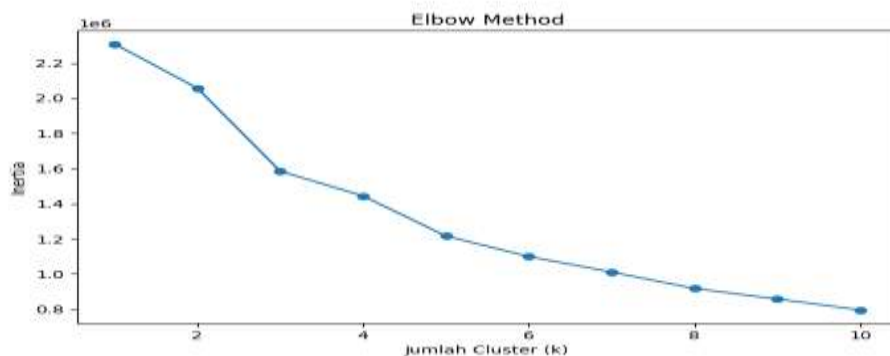
N o	Customer _Name	transactio n_count	total_ spent	avg_s pent	total_i tems	avg_i tems	discount _ratio	promo _ratio
1	Aaron Acevedo	4	278.95	69.63 7500	13	3.250 000	0.50000 0	0.7500 00
2	Aaron Acosta	8	252.32	31.54 1667	13	4.333 333	0.66666 7	0.6666 67
3	Aaron Adams	32	773.26	64.63 3333	77	4.666 667	0.33333 3	1.0000 00
4	Aaron Adkins	3	118.52	39.50 6667	11	3.666 667	0.00000 0	0.6666 67
5	Aaron Aguilar	5	238.76	47.75 2000	32	6.400 000	0.20000 0	1.0000 00

Tabel 2 memperlihatkan bahwa setiap pelanggan memiliki karakteristik transaksi yang berbeda. Perbedaan tersebut dapat diamati dari jumlah transaksi, nilai pengeluaran, jumlah barang yang dibeli, serta keterlibatan terhadap diskon dan promosi. Sebagai contoh, pelanggan dengan jumlah transaksi tinggi tidak selalu memiliki rata-rata pengeluaran atau rasio penggunaan diskon yang sama dengan pelanggan lainnya. Variasi karakteristik inilah yang menjadi dasar penggunaan *K-Means* untuk mengidentifikasi kelompok pelanggan yang memiliki pola transaksi relatif serupa.

Perubahan jumlah data dari 1.000.000 menjadi 329.738 tidak menunjukkan adanya pengurangan transaksi secara acak. Perubahan tersebut merupakan konsekuensi dari proses agregasi, karena sejumlah transaksi yang berasal dari pelanggan yang sama dirangkum menjadi satu representasi pelanggan. Dengan pendekatan tersebut, proses clustering dilakukan pada tingkat pelanggan (*customer level*), bukan pada tingkat transaksi (*transaction level*).

4.2 Hasil Segmentasi Pelanggan Menggunakan K-Means

Tahap berikutnya adalah melakukan segmentasi terhadap 329.738 data pelanggan menggunakan algoritma *K-Means*. Penentuan jumlah *cluster* dilakukan menggunakan *Elbow Method* dengan mengamati perubahan nilai *Within-Cluster Sum of Squares* (WCSS). Hasil segmentasi dengan *Elbow Method* dapat dilihat pada Gambar 1.



Gambar 1. Hasil Elbow Method untuk Penentuan Jumlah Cluster

Berdasarkan pola penurunan WCSS pada grafik *Elbow Method*, nilai $k = 3$ digunakan sebagai jumlah *cluster* dalam proses segmentasi. Pemilihan tersebut didasarkan pada titik perubahan ketika penurunan WCSS mulai menunjukkan kecenderungan yang lebih landai dibandingkan nilai k sebelumnya. Dengan demikian, data pelanggan dibagi menjadi tiga kelompok berdasarkan kedekatan karakteristik transaksi terhadap centroid masing-masing *cluster*.

Hasil proses *K-Means* menunjukkan distribusi anggota *cluster* yang tidak seimbang. Cluster 0 terdiri atas 188.607 pelanggan atau 57,2%, Cluster 1 sebanyak 3.171 pelanggan atau 0,96%, dan Cluster 2 sebanyak 137.960 pelanggan atau 41,84%. Jumlah anggota Cluster 1 yang relatif kecil juga menjadi temuan yang perlu diperhatikan. Hasil tersebut menunjukkan adanya sekelompok kecil pelanggan dengan karakteristik transaksi yang berbeda dari dua kelompok dominan lainnya. Meskipun demikian, jumlah anggota cluster saja belum cukup untuk memberikan interpretasi perilaku pelanggan. Karakterisasi lebih lanjut perlu didasarkan pada nilai atribut atau centroid masing-masing cluster agar pemberian label terhadap setiap kelompok dapat dilakukan secara objektif.

4.3 Hasil Analisis Pola Pembelian Menggunakan FP-Growth

Setelah segmentasi pelanggan dilakukan, FP-Growth diterapkan secara terpisah pada transaksi Kluster 0, Kluster 1, dan Kluster 2 untuk menemukan frequent itemsets serta membentuk aturan asosiasi. Pengujian menggunakan tiga skenario nilai minimum

support dan confidence guna mengetahui perubahan jumlah pola ketika nilai ambang diperketat. Skenario I menggunakan nilai 0,1% untuk kedua parameter, Skenario II menggunakan 0,2%, sedangkan Skenario III menetapkan minimum support 30% dan minimum confidence 50%. Hasil pengujian setiap skenario disajikan pada Tabel 3.

Tabel 3. Hasil Pengujian FP-Growth pada Setiap Cluster

Skenario	Cluster	Min. Support	Min. Confidence	Jumlah Transaksi	Frequent Itemset	Association Rules
I	0	0,1%	0,1%	602.911	81	6.478
I	1	0,1%	0,1%	155.711	81	6.312
I	2	0,1%	0,1%	241.378	81	6.432
II	0	0,2%	0,2%	602.911	81	160
II	1	0,2%	0,2%	155.711	81	160
II	2	0,2%	0,2%	241.378	81	160
III	0	30%	50%	602.911	—	—
III	1	30%	50%	155.711	—	—
III	2	30%	50%	241.378	—	—

Tabel 3 menunjukkan bahwa perubahan nilai minimum support dan confidence memengaruhi jumlah aturan yang terbentuk. Nilai ambang yang rendah pada Skenario I menghasilkan 6.478 aturan pada Kluster 0, 6.312 aturan pada Kluster 1, dan 6.432 aturan pada Kluster 2. Meskipun memberikan cakupan pola yang luas, jumlah tersebut relatif besar sehingga memerlukan penyaringan lebih lanjut agar aturan yang relevan lebih mudah diidentifikasi.

Peningkatan nilai ambang menjadi 0,2% pada Skenario II mengurangi hasil menjadi 160 aturan pada setiap kluster tanpa menghilangkan 81 frequent itemsets yang ditemukan. Sebaliknya, Skenario III tidak menghasilkan pola pada ketiga kluster karena tidak ada kombinasi produk yang mencapai minimum support 30% dan minimum confidence 50%. Temuan ini memperlihatkan bahwa semakin tinggi nilai ambang, semakin sedikit pola yang memenuhi kriteria; bahkan, nilai yang terlalu tinggi dapat menyebabkan tidak terbentuknya aturan asosiasi.

Berdasarkan hasil tersebut, Skenario II dipilih untuk analisis lanjutan karena menghasilkan jumlah aturan yang lebih mudah dikelola dan tetap mempertahankan

frequent itemsets. Pemilihan ini didasarkan pada keseimbangan praktis antara ketersediaan pola dan jumlah aturan, bukan sebagai parameter optimal yang berlaku secara umum. Penetapan nilai ambang tetap perlu disesuaikan dengan jumlah transaksi, keragaman produk, dan distribusi kemunculan item pada dataset yang dianalisis.

4.4 Pembahasan

Hasil penelitian memperlihatkan bahwa penggunaan K-Means sebelum FP-Growth memungkinkan proses analisis dilakukan pada dua tingkat yang berbeda. K-Means memberikan informasi mengenai pengelompokan pelanggan berdasarkan karakteristik transaksi, sedangkan FP-Growth digunakan untuk menggali hubungan antar produk pada transaksi pelanggan yang telah berada dalam kelompok tertentu. Dengan struktur tersebut, analisis pola pembelian tidak dilakukan langsung terhadap seluruh pelanggan sebagai satu populasi.

Segmentasi menghasilkan tiga cluster dengan ukuran yang berbeda secara cukup mencolok. Cluster 0 dan Cluster 2 mencakup sebagian besar pelanggan, sedangkan Cluster 1 hanya memiliki 3.171 anggota. Hasil ini mengindikasikan adanya heterogenitas karakteristik transaksi dalam dataset. Akan tetapi, interpretasi mengenai karakteristik masing-masing cluster perlu dilakukan secara hati-hati. Label seperti pelanggan dengan aktivitas rendah, pelanggan aktif, atau pelanggan yang sensitif terhadap promosi seharusnya ditetapkan berdasarkan profil centroid dan nilai rata-rata atribut pada setiap cluster, bukan semata-mata berdasarkan jumlah anggotanya.

Pada tahap FP-Growth, hasil eksperimen juga menunjukkan bahwa pemilihan nilai minimum *support* dan minimum *confidence* merupakan bagian yang menentukan banyaknya aturan yang diperoleh. Nilai ambang yang terlalu rendah menghasilkan ribuan aturan, sedangkan nilai yang terlalu tinggi dapat menghilangkan seluruh pola. Temuan ini menunjukkan pentingnya penentuan parameter yang sesuai dengan distribusi transaksi pada dataset.

Penggunaan Skenario II memberikan kompromi yang lebih praktis pada eksperimen yang dilakukan karena masih mempertahankan *frequent itemset* sekaligus mengurangi jumlah aturan dari lebih dari enam ribu menjadi 160 aturan pada setiap cluster. Meskipun demikian, jumlah aturan bukan satu-satunya indikator kualitas pola asosiasi. Analisis yang lebih mendalam masih diperlukan untuk mengetahui kekuatan dan

kegunaan setiap aturan, misalnya dengan mempertimbangkan nilai *lift* serta membandingkan aturan dominan yang terbentuk pada masing-masing cluster.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil pengujian, skenario II dengan nilai minimum *support* sebesar 0,2% dan minimum *confidence* sebesar 0,2% dipilih sebagai skenario yang paling sesuai untuk digunakan dalam analisis. Skenario I menghasilkan jumlah *association rules* yang relatif tinggi, yaitu 6.478 aturan pada *Cluster* 0, 6.312 aturan pada *Cluster* 1, dan 6.432 aturan pada *Cluster* 2, sehingga pola yang dihasilkan menjadi kurang selektif dan lebih sulit untuk dianalisis. Sementara itu, Skenario II menghasilkan 160 *association rules* pada masing-masing *cluster* dengan tetap menghasilkan *frequent itemset* sehingga jumlah aturan yang diperoleh lebih terkendali dan masih dapat digunakan untuk analisis pola pembelian. Adapun Skenario III dengan minimum *support* 30% dan minimum *confidence* 50% tidak menghasilkan *frequent itemset* maupun *association rules* pada seluruh *cluster* karena nilai ambang yang digunakan terlalu tinggi untuk karakteristik data transaksi penelitian.

Penelitian selanjutnya disarankan menggunakan metode evaluasi *clustering* seperti *Silhouette Score* dan *Elbow Method* untuk memperoleh jumlah *cluster* yang lebih optimal. Selain itu, penggunaan data transaksi nyata perusahaan ritel dapat meningkatkan relevansi hasil rekomendasi. Sistem juga dapat dikembangkan dengan visualisasi interaktif, dashboard analitik, rekomendasi real-time, serta aplikasi mobile agar lebih fleksibel dan mudah digunakan.

DAFTAR REFERENSI

- Annisa. (2026). *Algoritma K- Means Clustering: Pengertian, Fungsi Dan Cara Kerja*. <https://fikti.umsu.ac.id/Algoritma-K-Means-Clustering-Pengertian-Fungsi-Dan-Cara-Kerja/>
- Arisandi, F., Kristanto, & Juliane, C. (2023). Identifikasi Strategi Penjualan Dengan Pendekatan asosiasi FP-Growth Pada Perusahaan Ritel Berkah. *Teknik Informatika Dan Sistem Informasi*, 10(1), 421–431.
- Bunda, Y. P. (2020). Algoritma FP-Growth Untuk Menganalisa Pola Pembelian Oleh-Oleh (Studi Kasus Di Pusat Oleh-Oleh Ummi Aufa Hakim). *Riau Journal Of Computer Science*, 06(01), 34–44.
- Hairani, H., & Ximenes Guterres, J. (2024). Exploring Customer Purchasing Patterns: A

- Study Utilizing FP-Growth Algorithm On Supermarket Transaction Data. *International Journal Of Engineering And Computer Science Applications (IJECSA)*, 3(1), 33–42. <https://doi.org/10.30812/Ijecsa.V3i1.3874>
- Jadid, U. N. (2021). *ANALISIS KEBUTUHAN KONSUMEN SEBAGAI STRATEGI* Abstract : Abstrak : 1(1), 1–13.
- M. Rizky, A. Utama, R. U. Et Al. (2020). *Penerapan Algoritma Fp - Growth Untuk Penentuan Pola Pembelian Transaksi Penjualan Pada Toko Kgs Rizky Motor*. 25(1), 20–28.
- Makhfuzi Fahmi, Zumrotun Naimah, Novi Setiani, D. D. C. N. (2014). *Klasterisasi Judul Buku Dengan Menggunakan*. 1–4.
- Mardiana, N. A., & Windari, W. O. (2024). *G-Tech : Jurnal Teknologi Terapan*. *G-Tech : Jurnal Teknologi Terapan*, 8(1), 186–195.
- Nikmah, N., Balqis, A., Ainurrillah, V., Putri, D., Arfian, Z., & Ikaningtyas, M. (2025). Peran Teknologi Dalam Meningkatkan Manajemen Operasional Studi Pada Sektor Industri Dan Ritel. *Journal Of Technology And System Information*, 2(3), 9. <https://doi.org/10.47134/Jtsi.V2i3.4368>
- Rifda. (2023). *Peran Dan Tantangan Dalam Bisnis Retail: Menggali Potensi Dan Kesuksesan*. <https://voffice.co.id/blog/peran-tantangan-bisnis-retail/>
- Rohmah, A., Sembiring, F., & Erfina, A. (2021). *Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus : Smk Yaspim Gegerbitung)*. 290–298.
- Saraswati, I. D. A. I., Gunawan, I. M. A. O., & Putra, I. M. A. W. (2024). Analisis Keranjang Belanja Pada Data Ritel Non-Toko Menggunakan Algoritma FP-Growth. *Jutisi : Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 13(3), 1692–1704. <https://doi.org/10.35889/Jutisi.V13i3.2238>
- Setiyani, S., Aminah, A., Rilvani, E., Bangsa, U. P., & Bekasi, K. (2025). *PERAN DAN STRATEGI DATA MINING DALAM*. 3(8).
- Suhada, S., Ratag, D., Gunawan, G., Wintana, D., & Hidayatulloh, T. (2020). Penerapan Algoritma Fp-Growth Untuk Menentukan Pola Pembelian Konsumen Pada Ahass Cibadak. *Swabumi*, 8(2), 118–126. <https://doi.org/10.31294/Swabumi.V8i2.8077>
- Valencia, S., & Tisno Atmojo, W. (2024). Analisis Pola Pembelian Pada Data Penjualan Cannngopi Menggunakan Algoritma FP-Growth. *Jurnal Komtika (Komputasi Dan Informatika)*, 8(2), 214–224. <https://doi.org/10.31603/Komtika.V8i2.12672>