



Clustering Toko Ritel Berdasarkan Pola Penjualan Produk Menggunakan Algoritma K-Means

Riska Amalia Praptiwi¹, Yulya Muharmi², Dhella Amelia³

^{1,2,3} Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung

riskamaliatiwi93@fmipa.unila.ac.id d. yulya.muharmi@fmipa.unila.ac.id . dhellaaamelia@fmipa.unila.ac.id

Abstract

This study analyzes retail store sales patterns in Pekanbaru and performs data-driven segmentation using the K-Means algorithm. The dataset consists of 965 stores with numerical features derived from preprocessing that includes total sales, transaction frequency, average purchase value, and product variety. The clustering process produced three store segments: high-performance, medium-performance, and low-performance. Cluster quality was evaluated by using a Silhouette Score of 0.603 and a Davies-Bouldin Index of 0.763. These indicate coherent and well-separated groups. Further, PCA visualization confirmed the distinct separation among clusters. The novelty of this research lies in applying K-Means clustering to local retail transaction data in Pekanbaru with comprehensive cluster evaluation, providing segmentation that better represents regional consumer behavior. These findings can be utilized for product distribution strategies, targeted promotions and performance evaluation. Future research may incorporate geographic location and customer demographics to enhance segmentation precision.

Keywords: store segmentation, K-Means, cluster, retail sales, data analysis

Abstrak

Penelitian ini menganalisis pola penjualan toko ritel di Pekanbaru dan melakukan segmentasi berbasis data menggunakan algoritma K-Means. Dataset terdiri dari 965 toko dengan variabel numerik hasil *preprocessing*, meliputi total penjualan, frekuensi transaksi, rata-rata nilai pembelian, dan variasi produk. Hasil *clustering* menunjukkan tiga segmen toko: performa tinggi, menengah, dan rendah, dengan evaluasi kualitas menggunakan *Silhouette Score* sebesar 0.603 dan *Davies-Bouldin Index* sebesar 0.763. Visualisasi PCA memperkuat pemisahan antar cluster secara intuitif. Kebaruan penelitian ini terletak pada penerapan K-Means pada data ritel lokal Pekanbaru dengan evaluasi cluster komprehensif, sehingga menghasilkan segmentasi yang lebih representatif terhadap perilaku konsumen daerah. Temuan ini dapat dimanfaatkan untuk strategi distribusi produk, target promosi, dan evaluasi kinerja toko berbasis data. Penelitian selanjutnya dapat menambahkan variabel lokasi geografis dan demografi pelanggan untuk meningkatkan presisi segmentasi.

Kata kunci: segmentasi toko, K-Means, klaster, penjualan ritel, analisis data

© 2026 Author
Creative Commons Attribution 4.0 International License



1. Pendahuluan

Pertumbuhan industri ritel modern menghasilkan volume data transaksi yang semakin besar dan kompleks, sehingga diperlukan pendekatan analitik yang mampu mengidentifikasi pola pembelian secara efektif untuk mendukung pengambilan keputusan bisnis. Analisis segmentasi berbasis data menjadi penting terutama pada ritel wilayah Pekanbaru, di mana variasi karakteristik pelanggan dan jenis produk memerlukan strategi pemasaran yang lebih terarah. Berbagai penelitian terbaru menunjukkan bahwa teknik data mining, khususnya *clustering*, mampu membantu perusahaan memahami perilaku pelanggan dan meningkatkan efisiensi strategi bisnis melalui pengelompokan pola transaksi yang serupa [1].

Dalam lima tahun terakhir, metode K-Means dan variasinya banyak digunakan dalam studi segmentasi pelanggan ritel karena kemampuannya dalam mengolah data berskala besar dengan efisiensi komputasi yang baik. Studi internasional menunjukkan bahwa pengembangan algoritma *clustering* dapat meningkatkan kualitas segmentasi melalui optimasi centroid dan evaluasi metrik internal [2]. Penelitian lain di bidang retail analytics juga menekankan pentingnya preprocessing data dan pemilihan fitur transaksi untuk meningkatkan akurasi cluster yang dihasilkan [3][4]. Di Indonesia, beberapa penelitian nasional juga telah memanfaatkan *clustering* untuk analisis transaksi ritel, namun sebagian besar masih terbatas pada dataset skala kecil atau tidak berfokus pada konteks regional tertentu seperti Pekanbaru [5][6]. Perbedaan fokus tersebut menunjukkan adanya peluang penelitian yang lebih kontekstual dengan memanfaatkan data transaksi ritel lokal. Meskipun berbagai studi telah menunjukkan efektivitas *clustering* dalam segmentasi pelanggan, masih terdapat gap penelitian pada penerapan analisis *clustering* berbasis data transaksi ritel lokal yang memanfaatkan data aktual wilayah Pekanbaru. Sebagian penelitian internasional lebih banyak menggunakan dataset publik global, sehingga belum sepenuhnya merepresentasikan karakteristik perilaku konsumen di daerah berkembang. Selain itu, beberapa penelitian sebelumnya belum mengintegrasikan analisis evaluasi cluster secara komprehensif untuk mendukung rekomendasi strategi bisnis ritel berbasis data [7][8].

Berdasarkan gap tersebut, penelitian ini dilakukan untuk menganalisis pola transaksi ritel di wilayah Pekanbaru menggunakan pendekatan *clustering* guna menghasilkan segmentasi data yang lebih representatif terhadap perilaku pelanggan lokal. Untuk mendukung proses tersebut, penelitian ini memanfaatkan fitur numerik hasil ekstraksi data transaksi seperti total penjualan, frekuensi transaksi, rata-rata nilai pembelian, serta proporsi kategori produk yang dianalisis menggunakan algoritma K-Means. Beberapa penelitian sebelumnya menunjukkan bahwa *clustering* berbasis K-Means efektif dalam mengidentifikasi pola data dan meningkatkan kualitas segmentasi melalui evaluasi metrik internal [9][10]. Teknik asosiasi dan *clustering* juga mampu mengidentifikasi pola perilaku konsumen serta hubungan antar produk yang sering muncul dalam transaksi digital, sehingga dapat menjadi dasar strategi pemasaran berbasis data [11]. Selain itu, penerapan algoritma K-Means dalam penelitian lain menunjukkan bahwa metode *clustering* efektif digunakan untuk proses segmentasi data dan pemetaan kelompok berdasarkan karakteristik tertentu melalui proses iteratif centroid [12].

State of the art penelitian ini terletak pada penerapan *clustering* pada data transaksi ritel Pekanbaru dengan fokus pada analisis segmentasi berbasis wilayah serta evaluasi performa cluster untuk menghasilkan insight bisnis yang relevan. Adapun tujuan penelitian ini adalah: (1) mengidentifikasi pola transaksi ritel berdasarkan data penjualan, (2) membangun model *clustering* untuk segmentasi data ritel Pekanbaru, dan (3) mengevaluasi hasil segmentasi sebagai dasar rekomendasi pengambilan keputusan berbasis data.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis data mining untuk melakukan pengelompokan toko ritel di wilayah Pekanbaru berdasarkan pola penjualan produk.

2.1 Tahapan Penelitian

Tahapan penelitian ini disusun secara sistematis untuk memastikan proses *clustering* dapat dilakukan secara terstruktur dan dapat direplikasi. Alur penelitian ditunjukkan pada Gambar alur penelitian dan dijelaskan sebagai berikut:



Gambar 1. Tahapan Penelitian

1. Data Collection (Pengumpulan Data)

Tahap awal penelitian dilakukan dengan mengumpulkan data transaksi penjualan toko ritel di wilayah Pekanbaru. Dataset berisi informasi seperti kode toko, kategori produk, jumlah transaksi, tanggal transaksi, serta variabel pendukung lainnya yang relevan dengan pola penjualan.

2. Data Understanding & Selection (Pemahaman dan Seleksi Data)

Data yang diperoleh dianalisis untuk memahami karakteristik atribut serta menentukan fitur yang akan digunakan dalam proses *clustering*. Atribut yang tidak relevan atau memiliki nilai kosong berlebihan dieliminasi agar kualitas analisis meningkat.

3. Data Preprocessing (Pra-pemrosesan Data)

Proses ini diawali dengan pembersihan data melalui penghapusan duplikasi berdasarkan kombinasi atribut CUSTID dan tanggal transaksi, sehingga setiap transaksi yang digunakan bersifat unik dan tidak terjadi pengulangan pencatatan. Selanjutnya dilakukan penanganan missing value pada atribut utama, di mana data yang memiliki nilai kosong dihapus (*drop*) karena proporsinya relatif kecil (kurang dari 5% dari total dataset) sehingga tidak memberikan dampak signifikan terhadap distribusi data. Tahap berikutnya adalah transformasi data, yang meliputi penyesuaian tipe data numerik serta konversi format tanggal agar sesuai dengan kebutuhan proses analisis. Melalui tahapan ini, diperoleh dataset yang bersih, konsisten, dan siap digunakan pada proses *feature engineering* dan *clustering*. Normalisasi nilai menggunakan metode Min-Max Scaling agar setiap atribut memiliki skala yang sebanding [13]. Normalisasi dilakukan dengan rumus:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

di mana:

X adalah nilai asli, Xmin adalah nilai minimum dan Xmax adalah nilai maksimum.

Hasil normalisasi menghasilkan nilai dalam rentang 0 hingga 1. Proses ini penting untuk menghindari dominasi fitur tertentu dalam perhitungan jarak Euclidean pada algoritma K-Means.

4. Feature Engineering (Pembentukan Fitur Pola Penjualan)

Data transaksi diolah menjadi representasi numerik seperti total penjualan per kategori produk, frekuensi transaksi, dan rata-rata jumlah pembelian. Seluruh fitur dihitung berdasarkan agregasi data per toko dalam periode tahun 2023. Fitur yang digunakan meliputi:

a) Total Penjualan (Total Sales)

Total nilai transaksi per toko:

$$\text{Total Sales}_i = \sum_{j=1}^n \text{Transaksi}_{ij} \quad (2)$$

- b) Frekuensi Transaksi (Transaction Frequency)

Jumlah transaksi per toko:

$$Frequency_i = count(Transaksi_i) \quad (3)$$

- c) Rata-rata Nilai Transaksi (Average Purchase Value)

$$Average = \frac{Total\ Sales_i}{Frequency_i} \quad (4)$$

- d) Variasi Produk (Product Variety)

Jumlah kategori produk unik yang dijual:

$$Variety_i = count(Distinct\ Kategori_{produk_i}) \quad (5)$$

Fitur-fitur ini digunakan untuk membentuk profil numerik masing-masing toko ritel.

5. K Selection (Penentuan Jumlah Cluster)

Jumlah cluster optimal ditentukan menggunakan metode *Elbow* dengan menganalisis perubahan nilai *Within Cluster Sum of Squares* (WCSS) pada beberapa variasi nilai K. WCSS merupakan jumlah kuadrat jarak antara setiap data dengan centroid pada cluster-nya, yang dirumuskan sebagai:

$$WCSS = \sum_{k=1}^K \sum_{x_i \in C_k} (x_i - C_k)^2 \quad (6)$$

di mana x_i adalah data ke- i , C_k adalah centroid cluster ke- k , dan C_k adalah himpunan data dalam cluster tersebut.

Nilai K diuji dalam beberapa variasi untuk melihat perubahan nilai WCSS. Pendekatan ini digunakan untuk memperoleh jumlah cluster yang menghasilkan pemisahan data paling optimal dan menghindari *over-clustering*, sebagaimana direkomendasikan dalam studi validitas cluster [14].

6. Proses *Clustering* Menggunakan Algoritma K-Means

Dataset yang telah dipreproses kemudian diproses menggunakan algoritma K-Means untuk mengelompokkan toko ritel berdasarkan kemiripan pola penjualan. Algoritma ini bekerja melalui beberapa tahapan, yaitu:

- a) Inisialisasi centroid secara acak
- b) Menghitung jarak antara data dan centroid
- c) Mengelompokkan data ke cluster terdekat
- d) Memperbarui centroid berdasarkan rata-rata anggota cluster
- e) Mengulangi proses hingga centroid konvergen

Perhitungan jarak antar data menggunakan metode *Euclidean Distance* yang dirumuskan sebagai:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (7)$$

Implementasi algoritma dilakukan menggunakan bahasa pemrograman Python dengan pustaka Scikit learn. Parameter yang digunakan meliputi:

- a) `n_init`=sebagai jumlah inisialisasi centroid
- b) `max_iter`=300 sebagai batas maksimum iterasi
- c) metrik jarak Euclidean Distance

Pengaturan parameter ini bertujuan menjaga stabilitas hasil *clustering* dan memastikan proses konvergensi berjalan optimal [15].

7. Evaluasi Hasil *Clustering*

Hasil cluster dievaluasi menggunakan metrik internal seperti Silhouette Score dan Davies-Bouldin Index untuk mengukur kualitas pemisahan cluster serta konsistensi kelompok yang terbentuk. Silhouette Score digunakan untuk menilai tingkat kedekatan setiap data terhadap cluster-nya dibandingkan dengan cluster lain, sedangkan Davies-Bouldin Index digunakan untuk mengukur rasio kesamaan intra-cluster terhadap separasi antar cluster sehingga memberikan gambaran objektif terhadap kualitas segmentasi yang dihasilkan [16].

8. Analisis dan Interpretasi Cluster

Setiap cluster dianalisis untuk mengidentifikasi karakteristik toko ritel, misalnya cluster dengan penjualan tinggi, kategori produk dominan, atau frekuensi transaksi tertentu. Hasil interpretasi digunakan sebagai dasar rekomendasi strategi bisnis ritel.

9. Visualisasi dan Penyajian Hasil

Tahap akhir dilakukan visualisasi cluster menggunakan grafik atau diagram untuk mempermudah pemahaman pola segmentasi dan mendukung pengambilan keputusan berbasis data.

2.2 Objek dan Variabel Penelitian

Objek penelitian ini adalah toko ritel yang beroperasi di wilayah Pekanbaru dengan memanfaatkan data transaksi penjualan tahun 2023 sebagai dasar analisis. Data transaksi diolah menjadi representasi numerik untuk

menggambarkan karakteristik pola penjualan setiap toko sehingga mampu menghasilkan segmentasi yang lebih relevan terhadap kebutuhan pengambilan keputusan bisnis ritel.

Variabel penelitian difokuskan pada fitur hasil pengolahan data yang merepresentasikan perilaku penjualan, bukan pada atribut mentah transaksi. Pemilihan variabel disesuaikan dengan kebutuhan algoritma *clustering* yang memerlukan representasi numerik yang konsisten dan mampu membedakan karakteristik antar toko. Variabel yang digunakan meliputi: (1) total nilai penjualan sebagai indikator performa penjualan, (2) frekuensi transaksi untuk menggambarkan tingkat aktivitas toko, (3) rata-rata nilai transaksi untuk melihat kecenderungan volume pembelian, (4) distribusi penjualan berdasarkan kategori produk sebagai representasi pola preferensi penjualan, serta (5) indeks pola penjualan toko sebagai fitur agregasi yang membentuk profil numerik tiap toko ritel. Seluruh variabel kemudian dinormalisasi agar berada pada skala yang sebanding sebelum digunakan dalam proses *clustering* menggunakan algoritma K-Means. Pendekatan berbasis fitur agregasi ini mendukung pembentukan segmentasi yang lebih informatif dan kontekstual terhadap kebutuhan analisis bisnis ritel.

3. Hasil dan Pembahasan

3.1 Hasil Pra-pemrosesan dan Pembentukan Fitur

Tahap awal pengolahan data dilakukan melalui proses *preprocessing* terhadap data transaksi penjualan toko ritel wilayah Pekanbaru tahun 2023. Proses ini meliputi pembersihan data, transformasi numerik, serta normalisasi nilai atribut menggunakan metode Min-Max Scaling untuk memastikan setiap variabel memiliki kontribusi yang seimbang dalam proses *clustering*. Dataset yang digunakan terdiri dari 965 data toko ritel yang telah melalui proses *preprocessing* sebelum dilakukan *clustering*. Normalisasi dilakukan menggunakan (1) dimana proses ini menghasilkan nilai dalam rentang 0 hingga 1 sehingga dapat menghindari dominasi variabel tertentu dalam perhitungan jarak Euclidean pada algoritma K-Means. Pada toko dengan CUSTID 5100000075 diperoleh total penjualan sebesar Rp121.000.000 (nilai sebelum normalisasi) yang merupakan hasil penjumlahan seluruh transaksi pada toko tersebut. Nilai ini kemudian dinormalisasi menggunakan metode Min-Max Scaling dengan mempertimbangkan nilai minimum dan maksimum pada seluruh dataset.

Dengan nilai minimum dan maksimum yang diperoleh dari dataset sebesar Rp10.000 dan nilai maksimum sebesar Rp2.300.000.000, maka perhitungan normalisasi dilakukan sebagai berikut:

$$X' = \frac{121.000.000 - 10.000}{2.300.000.000 - 10.000} = 0,052810$$

Nilai tersebut sesuai dengan hasil normalisasi yang ditampilkan pada Tabel 1, sehingga menunjukkan bahwa proses normalisasi telah dilakukan secara konsisten pada seluruh data.

Data transaksi selanjutnya diolah melalui proses *feature engineering* untuk membentuk fitur agregasi yang merepresentasikan pola penjualan masing-masing toko. Proses ini dilakukan berdasarkan perhitungan total penjualan, frekuensi transaksi, rata-rata nilai transaksi, serta variasi produk sebagaimana dirumuskan pada Persamaan (2) hingga (5). STotal nilai penjualan dihitung sebagai jumlah seluruh nilai transaksi pada setiap toko menggunakan fungsi penjumlahan (*sum*), frekuensi transaksi dihitung sebagai jumlah nota transaksi menggunakan fungsi *count*, rata-rata nilai transaksi diperoleh dari pembagian total penjualan dengan frekuensi transaksi, sedangkan variasi produk dihitung sebagai jumlah produk unik menggunakan fungsi *unique*. Proses ini bertujuan mengubah data transaksi mentah menjadi representasi numerik yang lebih informatif sehingga dapat meningkatkan kualitas segmentasi yang dihasilkan oleh algoritma K-Means. Hasil *preprocessing* menunjukkan bahwa seluruh variabel telah berada pada rentang nilai yang sebanding, yaitu antara 0 hingga 1. Kondisi ini penting untuk menghindari dominasi skala tertentu pada perhitungan jarak Euclidean. Contoh hasil transformasi data setelah proses normalisasi ditunjukkan pada Tabel 1, yang memperlihatkan representasi numerik fitur yang digunakan sebagai input algoritma *clustering*.

Tabel 1. Hasil Normalisasi Fitur Penjualan Toko Ritel

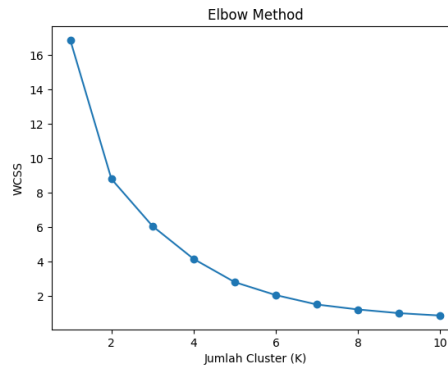
No.	CUSTID	total_penjualan	frekuensi_transaksi	variasi_produk
1	5100000075	0.052810	0.182524	0.30
2	5100000093	0.000266	0.005825	0.10
3	5100000210	0.000410	0.007767	0.10
4	5100000241	0.000435	0.005825	0.15
5	5100000253	0.000329	0.011650	0.05

Catatan: Tabel ini hanya menampilkan sebagian contoh data hasil normalisasi. Seluruh data transaksi telah diproses dan digunakan dalam tahap *clustering*.

3.2 Penentuan Jumlah Cluster Optimal

Setelah proses preprocessing dan normalisasi selesai dilakukan, tahap berikutnya adalah menentukan jumlah cluster optimal yang akan digunakan pada algoritma K-Means. Penentuan jumlah cluster merupakan langkah penting karena akan mempengaruhi kualitas segmentasi data dan interpretasi pola penjualan toko ritel.

Pada penelitian ini, jumlah cluster ditentukan menggunakan metode Elbow dengan menghitung nilai *Within Cluster Sum of Squares* (WCSS) pada beberapa variasi nilai K. Nilai WCSS dihitung menggunakan (6) dan dianalisis pada berbagai nilai K untuk menentukan jumlah cluster optimal.



Gambar 2. Grafik Elbow Method untuk Penentuan Jumlah Cluster Optimal

Berdasarkan hasil pengujian terhadap nilai K dari 1 hingga 10, terlihat bahwa penurunan nilai WCSS terjadi secara signifikan hingga K=3, kemudian cenderung melandai pada nilai K berikutnya. Kondisi ini menunjukkan bahwa penambahan cluster setelah K=3 tidak memberikan peningkatan pemisahan data yang berarti. Oleh karena itu, jumlah cluster optimal yang digunakan pada penelitian ini adalah K=3.

3.3 Hasil Proses Clustering Menggunakan K-Means

Setelah jumlah cluster optimal ditentukan, proses *clustering* dilakukan menggunakan algoritma K-Means terhadap data hasil normalisasi. Proses ini menghasilkan tiga kelompok toko ritel yang memiliki karakteristik pola penjualan yang berbeda. Perhitungan jarak antar data dilakukan menggunakan Euclidean Distance sesuai dengan (7).

Distribusi jumlah anggota pada setiap cluster ditunjukkan pada Tabel 2. Hasil tersebut memperlihatkan bahwa sebagian besar toko berada pada cluster 1, sedangkan cluster 2 memiliki jumlah anggota yang lebih sedikit dan menunjukkan karakteristik khusus dibandingkan cluster lainnya.

Tabel 2. Distribusi Jumlah Toko pada Setiap Cluster

Cluster	Jumlah Toko
0	291
1	649
2	25

Untuk memahami karakteristik setiap kelompok, dilakukan analisis rata-rata nilai fitur pada masing-masing cluster sebagaimana ditampilkan pada Tabel 3. Nilai rata-rata ini memberikan gambaran profil penjualan yang membedakan tiap cluster.

Tabel 3. Rata-rata Fitur pada Setiap Cluster

Cluster	total penjualan	frekuensi transaksi	variasi produk
0	0.065793	0.138471	0.462121
1	0.001169	0.008408	0.047458
2	0.001164	0.015749	0.206537

Berdasarkan nilai rata-rata fitur tersebut, karakteristik setiap cluster dapat diinterpretasikan sebagai berikut:

1. **Cluster 0 (Performa Tinggi)**

Cluster ini menunjukkan nilai total penjualan, frekuensi transaksi, dan variasi produk tertinggi dibandingkan cluster lainnya. Kelompok ini merepresentasikan segmen toko dengan performa penjualan tinggi dan aktivitas transaksi yang intensif.

2. **Cluster 1 (Performa Rendah)**

Cluster ini memiliki nilai rata-rata terendah pada seluruh fitur, yang menggambarkan toko dengan aktivitas transaksi rendah serta variasi produk yang terbatas. Cluster ini merepresentasikan segmen toko dengan performa penjualan rendah namun relatif stabil.

3. Cluster 2 (Performa Menengah)

Cluster ini berada pada kategori menengah, dengan frekuensi transaksi dan variasi produk lebih tinggi dibandingkan Cluster 1, namun masih berada di bawah Cluster 0. Kelompok ini menunjukkan karakteristik toko dengan aktivitas penjualan moderat.

Perbedaan karakteristik antar cluster menunjukkan bahwa algoritma K-Means berhasil memisahkan data toko ritel ke dalam segmen dengan profil penjualan yang berbeda secara jelas. Hasil segmentasi ini dapat digunakan sebagai dasar dalam analisis strategi distribusi produk, penentuan target promosi, serta evaluasi kinerja toko secara lebih terarah.

3.4 Evaluasi Kualitas Cluster

Kualitas cluster dievaluasi menggunakan metrik internal, yaitu Silhouette Score dan Davies–Bouldin Index, untuk menilai sejauh mana pemisahan antar cluster dan konsistensi internal setiap cluster. Silhouette Score mengukur sejauh mana objek lebih mirip dengan anggota cluster-nya sendiri dibandingkan dengan cluster lainnya, sedangkan Davies–Bouldin Index menilai rasio jarak intra-cluster terhadap jarak inter-cluster, sehingga memberikan indikasi tingkat tumpang tindih antar cluster. Nilai Silhouette Score dihitung untuk mengevaluasi kualitas pemisahan cluster. Hasil evaluasi menunjukkan bahwa model *clustering* menghasilkan nilai Silhouette Score sebesar 0.603 dan Davies–Bouldin Index sebesar 0.763. Nilai Silhouette Score yang mendekati 1 menunjukkan bahwa sebagian besar data telah terkelompok dengan baik sesuai karakteristik cluster-nya, sedangkan nilai Davies–Bouldin Index yang relatif rendah mengindikasikan bahwa cluster yang terbentuk memiliki separasi yang memadai dan tingkat homogenitas internal yang baik. Dengan demikian, konfigurasi tiga cluster (Cluster Tinggi, Menengah, dan Rendah) dinilai representatif terhadap struktur data penjualan toko ritel di wilayah Pekanbaru.

3.5 Analisis dan Interpretasi Segmentasi Toko Ritel

Hasil *clustering* menggunakan algoritma K-Means menghasilkan tiga segmen toko ritel dengan karakteristik pola penjualan yang berbeda berdasarkan fitur total penjualan, frekuensi transaksi, dan variasi produk. Perbedaan nilai rata-rata fitur pada setiap cluster menunjukkan adanya pola perilaku penjualan yang khas dan dapat digunakan untuk memahami profil toko secara lebih objektif. Penamaan segmen tidak didasarkan pada nomor cluster, melainkan pada perbandingan nilai rata-rata fitur penjualan pada masing-masing cluster.

Tabel 4. Interpretasi Karakteristik Segmentasi Toko Ritel

Cluster	Segmen	Karakteristik Utama	Interpretasi
0	Cluster Tinggi	Total penjualan tinggi, frekuensi transaksi tinggi, variasi produk luas	Toko dengan performa penjualan tinggi dan prioritas strategi distribusi/promosi
2	Cluster Menengah	Nilai penjualan dan frekuensi transaksi menengah	Toko dengan aktivitas penjualan moderat dan potensi pengembangan
1	Cluster Rendah	Total penjualan rendah, frekuensi transaksi rendah, variasi produk terbatas	Toko dengan performa penjualan rendah yang memerlukan evaluasi strategi

Berdasarkan hasil *clustering* dan evaluasi kualitas cluster, segmentasi toko ritel dibagi menjadi tiga kelompok utama, yaitu Cluster Tinggi, Cluster Menengah, dan Cluster Rendah. Interpretasi karakteristik masing-masing segmen disajikan pada Tabel 4.

1. Cluster Tinggi (Cluster 0) merepresentasikan toko dengan performa penjualan tinggi, ditandai oleh total penjualan besar, frekuensi transaksi tinggi, serta variasi produk yang luas. Toko pada segmen ini berpotensi menjadi prioritas dalam strategi distribusi produk dan promosi.
2. Cluster Menengah (Cluster 2) mencerminkan toko dengan aktivitas penjualan moderat. Meskipun performanya belum setinggi Cluster Tinggi, segmen ini memiliki potensi untuk dikembangkan melalui strategi pemasaran dan optimalisasi variasi produk.
3. Cluster Rendah (Cluster 1) menunjukkan toko dengan aktivitas penjualan rendah dan variasi produk terbatas. Segmen ini memerlukan evaluasi strategi bisnis, seperti penyesuaian produk atau pendekatan promosi yang lebih terarah.

Untuk meningkatkan nilai praktis dari hasil segmentasi, interpretasi karakteristik masing-masing cluster tidak hanya digunakan untuk memahami profil toko ritel, tetapi juga sebagai dasar dalam perumusan strategi bisnis yang lebih terarah. Oleh karena itu, setiap segmen cluster dapat dikaitkan dengan contoh strategi operasional dan pemasaran yang relevan sesuai dengan karakteristik performa penjualannya.

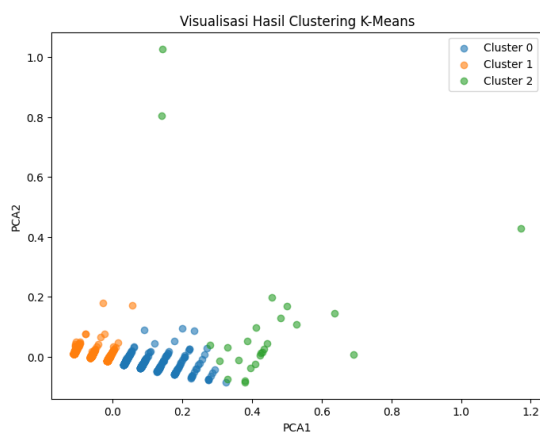
Tabel 5. Contoh Strategi Bisnis Berdasarkan Segmentasi Toko Ritel

Segmen Cluster	Karakteristik Umum	Contoh Strategi Nyata
Cluster Tinggi (Cluster 0)	Performa penjualan tinggi	Distribusi produk premium, promosi loyalitas pelanggan
Cluster Menengah (Cluster 2)	Aktivitas penjualan moderat	Diversifikasi produk, peningkatan frekuensi transaksi
Cluster Rendah (Cluster 1)	Penjualan dan transaksi rendah	Strategi diskon selektif atau evaluasi efisiensi operasional

Berdasarkan Tabel 5, toko yang termasuk dalam Cluster Tinggi dapat diprioritaskan dalam distribusi produk premium dan program loyalitas pelanggan. Cluster Menengah menunjukkan potensi pengembangan melalui diversifikasi produk dan peningkatan frekuensi transaksi, sedangkan Cluster Rendah memerlukan pendekatan strategis berupa promosi harga atau evaluasi efisiensi operasional. Pendekatan ini menunjukkan bahwa hasil *clustering* dapat dimanfaatkan secara langsung sebagai dasar pengambilan keputusan strategis di sektor ritel.

3.6 Visualisasi Hasil *Clustering*

Visualisasi hasil *clustering* ditampilkan dalam proyeksi dua dimensi menggunakan PCA, seperti terlihat pada Gambar 3. Cluster ditandai dengan warna berbeda untuk mempermudah identifikasi pola segmentasi.



Gambar 3. Visualisasi Hasil *Clustering*

Hasil visualisasi menunjukkan bahwa ketiga cluster membentuk kelompok yang relatif terpisah, meskipun terdapat kedekatan antara beberapa cluster. Cluster 0 (biru) dan Cluster 1 (oranye) terlihat berada pada area yang lebih berdekatan, sedangkan Cluster 2 (hijau) cenderung menempati area yang relatif terpisah dalam ruang dua dimensi hasil reduksi PCA. Pemisahan ini menunjukkan adanya perbedaan karakteristik pola penjualan antar cluster, namun tidak secara langsung merepresentasikan tingkat performa penjualan.

Interpretasi tingkat performa toko ditentukan berdasarkan analisis nilai rata-rata fitur pada masing-masing cluster, bukan berdasarkan posisi visual semata. Hasil visualisasi ini memperkuat temuan evaluasi metrik internal, yaitu Silhouette Score sebesar 0.603 dan Davies–Bouldin Index sebesar 0.763, yang menunjukkan bahwa cluster yang terbentuk memiliki separasi yang cukup baik serta homogenitas internal yang memadai.

Dengan demikian, visualisasi PCA berfungsi sebagai alat bantu untuk memahami pola distribusi dan pemisahan cluster secara intuitif, serta mendukung analisis segmentasi toko ritel berbasis data sebagai dasar pengambilan keputusan strategis, seperti penentuan distribusi produk, target promosi, dan evaluasi kinerja toko.

4. Kesimpulan

Berdasarkan hasil analisis *clustering* menggunakan metode K-Means, toko ritel dapat dibagi menjadi tiga segmen dengan karakteristik penjualan yang berbeda, yaitu berdasarkan frekuensi transaksi, variasi produk, dan total penjualan. Segmentasi ini menunjukkan adanya pola perilaku penjualan pada setiap toko, sehingga memungkinkan identifikasi kelompok toko dengan performa tinggi, menengah, dan rendah. Cluster 0 merepresentasikan toko dengan performa penjualan tinggi yang ditandai oleh nilai total penjualan, frekuensi transaksi, dan variasi produk yang lebih tinggi dibandingkan cluster lainnya. Cluster 1 mencerminkan toko dengan aktivitas penjualan rendah, ditunjukkan oleh frekuensi transaksi yang rendah serta variasi produk yang terbatas. Sementara itu, Cluster 2 berada pada kategori menengah, dengan karakteristik frekuensi transaksi dan variasi produk yang lebih tinggi dibandingkan Cluster 1, namun masih berada di bawah Cluster 0.

Evaluasi kualitas cluster menggunakan Silhouette Score sebesar 0.603 dan Davies–Bouldin Index sebesar 0.763 menunjukkan bahwa hasil segmentasi memiliki tingkat pemisahan antar cluster yang cukup baik serta konsistensi internal yang memadai. Visualisasi menggunakan Principal Component Analysis (PCA) juga menunjukkan distribusi cluster yang relatif terpisah secara visual, sehingga mendukung interpretasi hasil *clustering*, meskipun penentuan tingkat performa tetap didasarkan pada analisis nilai rata-rata fitur. Secara praktis, hasil segmentasi ini dapat dimanfaatkan untuk mendukung strategi distribusi produk, penentuan target promosi, serta evaluasi kinerja toko secara lebih objektif. Penelitian selanjutnya disarankan untuk menambahkan variabel lain seperti lokasi geografis dan karakteristik pelanggan guna meningkatkan akurasi dan kedalaman analisis segmentasi.

Daftar Rujukan

- [1] V. et al. Kumar, "Customer Segmentation Using Machine Learning in Retail Industry," *Appl. Sci.*, vol. 11, no. 17, 2021, doi: 10.3390/app11178162.
- [2] K. P. Sinaga and M.-S. Yang, "Unsupervised K-Means Clustering Algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [3] T. et al. Chen, "Retail Data Analytics and Customer Behaviour Modelling Using Clustering Techniques," *Expert Syst. Appl.*, vol. 187, no. 1, 2022, doi: 10.1016/j.eswa.2021.115972.
- [4] X. Li and H. Wang, "Improved K-Means for Large Scale Retail Data Segmentation," *Knowledge-Based Syst.*, vol. 261, no. 1, 2023, doi: 10.1016/j.knosys.2022.110182.
- [5] A. et al. Putra, "Penerapan Data Mining untuk Segmentasi Pelanggan Menggunakan K-Means," *J. RESTI*, vol. 4, no. 5, 2020, doi: 10.29207/resti.v4i5.2450.
- [6] D. et al. Rahmawati, "Analisis Clustering Data Penjualan Retail Menggunakan Algoritma K-Means," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 3, 2021, doi: 10.25126/jtiik.202183450.
- [7] Y. et al. Zhang, "Data Mining Techniques for Retail Market Basket Analysis," *IEEE Access*, vol. 8, no. 1, 2020, doi: 10.1109/ACCESS.2020.3001930.
- [8] J. et al. Silva, "Machine Learning-Based Customer Behaviour Analysis in Smart Retail," *Sensors*, vol. 24, no. 2, 2024, doi: 10.3390/s24020521.
- [9] Y. K. Febrina, R. Saputra, and K. Flomina, "Segmentasi Pelanggan Toko Hanifah Berdasarkan Analisis RFM dengan Metode K-Means Clustering," *J. Pustaka AI*, vol. 5, no. 2, pp. 274–282, 2025, doi: 10.55382/jurnalpustakaai.v5i2.1084.
- [10] D. P. Sari, W. Buana, and M. F. M. Sari, "Implementasi Data Mining pada Penjualan Barang dengan Teknik K Means," *J. Pustaka AI*, vol. 5, no. 1, 2025, doi: 10.55382/jurnalpustakaai.v5i1.955.
- [11] Y. Muharmi, N. Azwanti, S. T. Siska, and D. Amelia, "Association Rule Analysis of Student Consumptive Behavior on Shopee Twin Date Promo," *J. Sci. Appl. Informatics*, vol. 7, no. 3, 2024, [Online]. Available: <https://gas3kglangka.umb.ac.id/index.php/JSIAI/article/view/7390>
- [12] N. Azwanti and N. E. Putri, "Penerapan Algoritma K-Means untuk Pemetaan Penerimaan Bantuan Kesejahteraan Masyarakat di Kota Batam," in *Seminar Nasional Ilmu Sosial dan Teknologi (SNISTEK) 5*, 2023. [Online]. Available: <https://pdfs.semanticscholar.org/1d84/b6544555a5d8d7e593699a10d41bdf21ecf2.pdf>
- [13] M. Ahmed, R. Seraj, and S. Islam, "Customer Segmentation Using K-Means Clustering," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 5, 2020, doi: 10.14569/IJACSA.2020.0110573.
- [14] O. Arbelaitz, I. Gurrutxaga, and J. Muguerza, "An extensive comparative study of cluster validity indices," *Pattern Recognit.*, vol. 46, no. 1, 2013, doi: 10.1016/j.patcog.2012.07.021.
- [15] T. Hidayat and B. Surarso, "Penerapan Algoritma K-Means untuk Segmentasi Penjualan," *J. RESTI*, vol. 5, no. 2, 2021, doi: 10.29207/resti.v5i2.3001.
- [16] R. Sari and A. Wibowo, "Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 4, 2022, doi: 10.25126/jtiik.202294567.