



# Implementasi K-Means Clustering Berbasis RapidMiner untuk Optimalisasi Segmentasi Penjualan Produk dalam Meningkatkan Efektivitas Strategi Pemasaran

**Sufajar Butsianto\*, Arif Siswandi**

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pelita Bangsa, Bekasi  
Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Selatan, Bekasi, Jawa Barat, Indonesia

Email: <sup>1,\*</sup>sufajar@pelitabangsa.ac.id, <sup>2</sup>arif.siswandi@pelitabangsa.ac.id

Email Penulis Korespondensi: sufajar@pelitabangsa.ac.id

Submitted: 30/09/2025; Accepted: 31/10/2025; Published: 31/10/2025

**Abstrak**—Industri ritel elektronik di Indonesia mengalami pertumbuhan pesat seiring transformasi digital. Namun, data penjualan yang tersedia sering kali hanya tersimpan sebagai catatan transaksi tanpa analisis lebih lanjut, sehingga belum dimanfaatkan secara optimal untuk pengambilan keputusan pemasaran maupun segmentasi pelanggan. Penelitian ini bertujuan mengimplementasikan algoritma K-Means Clustering berbasis RapidMiner untuk menganalisis pola segmentasi produk elektronik di Toko XYZ. Dataset yang digunakan mencakup variabel Transaction\_ID, Product\_ID, Product\_Name, Category, Quantity, Unit\_Price, Revenue, dan Recency. Tahap penelitian meliputi pengumpulan data, preprocessing (filter, agregasi, dan normalisasi Z-Score), penerapan K-Means, hingga interpretasi hasil clustering. Penentuan jumlah cluster dalam penelitian ini menggunakan Elbow Method, yang menunjukkan titik optimal pada K=3, sehingga jumlah cluster tersebut dipilih untuk proses pengelompokan data. Berdasarkan hasil penerapan algoritma K-Means dengan tiga cluster tersebut, diperoleh: (1) Cluster 0 (High Sales & High Revenue) terdiri dari Smartphone, Laptop, dan Tablet sebagai produk unggulan dengan kontribusi hampir 60% total revenue; (2) Cluster 1 (Medium Sales & Moderate Revenue) mencakup Televisi, Kulkas, dan Smartwatch dengan kontribusi stabil sekitar 27%; dan (3) Cluster 2 (Low Sales & Low Revenue) berisi Mesin Cuci, Speaker, Headphone, dan Printer dengan kontribusi rendah, hanya 14%. Temuan ini memberikan dasar strategis bagi manajemen dalam pengambilan keputusan bisnis, seperti prioritas pengadaan, promosi musiman, bundling produk, hingga strategi clearance. Penelitian ini membuktikan bahwa penerapan data mining dengan K-Means Clustering efektif meningkatkan efisiensi operasional dan mendukung daya saing bisnis ritel elektronik di Indonesia..

**Kata kunci:** K-Means Clustering; Ritel Elektronik; Segmentasi Penjualan; Optimasi Strategi Pemasaran.

**Abstract**—The Indonesian electronic retail industry is experiencing rapid growth along with digital transformation. However, available sales data is often only stored as transaction records without further analysis, so it has not been optimally utilized for marketing decision making or customer segmentation. This study aims to implement the RapidMiner-based K-Means Clustering algorithm to analyze segmentation patterns of electronic products at XYZ Store. The dataset used includes the variables Transaction\_ID, Product\_ID, Product\_Name, Category, Quantity, Unit\_Price, Revenue, and Recency. The research stages include data collection, preprocessing (filtering, aggregation, and Z-Score normalization), K-Means application, and interpretation of clustering results. Determination of the number of clusters in this study uses the Elbow Method, which shows an optimal point at K = 3, so that number of clusters is chosen for the data grouping process. Based on the results of the application of the K-Means algorithm with the three clusters, the following are obtained: (1) Cluster 0 (High Sales & High Revenue) consisting of Smartphones, Laptops, and Tablets as superior products with a contribution of almost 60% of total revenue; (2) Cluster 1 (Medium Sales & Moderate Revenue) includes Televisions, Refrigerators, and Smartwatches with a stable contribution of around 27%; and (3) Cluster 2 (Low Sales & Low Revenue) contains Washing Machines, Speakers, Headphones, and Printers with a low contribution of only 14%. These findings provide a strategic basis for management in making business decisions, such as procurement priorities, seasonal promotions, product bundling, and clearance strategies. This study proves that the application of data mining with K-Means Clustering is effective in increasing operational efficiency and supporting the competitiveness of the electronics retail business in Indonesia.

**Keywords:** K-Means Clustering; Electronic Retail; Sales Segmentation; Marketing Strategy Optimization.

## 1. PENDAHULUAN

Industri ritel di Indonesia[1] dalam beberapa tahun terakhir menunjukkan perkembangan yang sangat pesat, terutama didorong oleh transformasi digital dan meningkatnya daya beli masyarakat. Data Bank Indonesia melalui Survei Penjualan Eceran mencatat bahwa Indeks Penjualan Riil (IPR) pada Juni 2025 tumbuh sebesar 2,0% year-on-year (yoY), lebih tinggi dibandingkan bulan sebelumnya yang hanya 1,9% yoY. Sementara itu, Trading Economics melaporkan penjualan ritel Indonesia meningkat 4,7% yoY pada Juli 2025, melonjak signifikan dibandingkan pertumbuhan 1,3% yoY pada bulan Juni. Peningkatan ini menandakan bahwa sektor ritel terus bertumbuh, dengan jumlah transaksi dan volume data penjualan yang semakin besar dari waktu ke waktu. Namun, di balik pertumbuhan tersebut, terdapat tantangan besar yang harus dihadapi[2]. Data transaksi yang melimpah sering kali tidak dimanfaatkan secara optimal karena hanya disimpan sebagai catatan penjualan tanpa analisis lebih lanjut. Padahal, data penjualan merupakan aset penting yang menyimpan informasi berharga mengenai perilaku konsumen[3], tren pembelian, hingga preferensi produk. Tanpa adanya proses analisis yang tepat, data mentah tersebut tidak akan memberikan nilai tambah strategis dan hanya menjadi tumpukan angka yang pasif. Kondisi ini semakin kompleks dengan adanya risiko shrinkage atau penyusutan persediaan yang terjadi akibat pencurian, kerusakan, kesalahan pencatatan, maupun faktor lain. Berbagai studi memperkirakan bahwa shrinkage dapat



mengurangi sekitar 1–2% dari total penjualan ritel, jumlah yang cukup signifikan bila dihitung dalam skala besar. Salah satu sektor ritel yang mengalami pertumbuhan cukup signifikan adalah penjualan produk elektronik, termasuk smartphone, laptop, dan perangkat rumah tangga. Meskipun demikian, Toko XYZ menghadapi tantangan dalam mengelola data penjualannya yang terus bertambah, terutama untuk mengetahui produk elektronik mana yang paling diminati konsumen, produk mana yang penjualannya stagnan, serta bagaimana menyusun strategi pengadaan dan promosi secara lebih tepat sasaran. Tanpa analisis yang memadai, data penjualan yang dimiliki hanya akan menjadi catatan transaksi tanpa memberikan nilai tambah bagi pengambilan keputusan bisnis.

Dalam menghadapi tantangan tersebut, data mining hadir sebagai solusi untuk menggali pola tersembunyi dalam data penjualan, termasuk pada penjualan produk elektronik di Toko XYZ. Salah satu metode yang relevan adalah K-Means Clustering, yang dapat mengelompokkan produk berdasarkan kemiripan pola penjualan sehingga memudahkan manajemen dalam membuat segmentasi produk elektronik. Dengan bantuan perangkat lunak RapidMiner, implementasi K-Means dapat dilakukan secara lebih praktis, cepat, dan efisien, serta menghasilkan visualisasi yang memudahkan interpretasi hasil analisis. Melalui teknik ini, data mentah yang semula hanya berupa catatan transaksi dapat diubah menjadi informasi yang bernilai strategis, misalnya dalam bentuk pola pembelian konsumen, tren produk terlaris[4], hingga kecenderungan perilaku pelanggan[5] dari waktu ke waktu. Hasil analisis semacam ini tidak hanya berguna untuk memahami kondisi pasar, tetapi juga dapat menjadi dasar dalam penyusunan strategi pemasaran, pengadaan barang, maupun penentuan harga yang lebih kompetitif. Salah satu pendekatan populer dalam data mining adalah clustering, yaitu proses pengelompokan data[6] ke dalam beberapa segmen berdasarkan kesamaan karakteristik tertentu. Dalam konteks ritel, clustering[7] memungkinkan pengelompokan produk atau konsumen ke dalam kategori-kategori spesifik, seperti produk dengan penjualan[8] tinggi, sedang, atau rendah, serta konsumen dengan frekuensi pembelian tertentu. Dengan adanya segmentasi ini, perusahaan dapat menyusun strategi promosi yang lebih terarah, menyesuaikan pengadaan barang sesuai kebutuhan pasar, dan mengelola stok secara lebih efisien untuk meminimalkan risiko kelebihan atau kekurangan persediaan. Di antara berbagai algoritma clustering, K-Means Clustering[9] merupakan salah satu metode yang paling banyak digunakan dalam analisis data karena kesederhanaannya dalam mengelompokkan data berdasarkan kemiripan karakteristik. Algoritma K-Means dipilih dalam penelitian ini karena mampu melakukan pengelompokan data secara efisien pada dataset berukuran besar dan menghasilkan cluster yang terstruktur serta mudah dianalisis. Dibandingkan metode clustering lain seperti Hierarchical Clustering yang kurang efisien untuk data besar atau DBSCAN yang lebih cocok untuk pola cluster tidak beraturan, K-Means lebih tepat untuk menghasilkan segmentasi produk dengan jumlah cluster yang jelas dan terdefinisi. Keunggulan ini menjadikan K-Means metode yang sesuai untuk tujuan analisis segmentasi penjualan dalam penelitian ini. K-Means Clustering[10] merupakan algoritma yang memiliki keunggulan berupa kesederhanaan konsep, kecepatan proses, serta efektivitas dalam mengolah data berukuran besar. Dengan prinsip pembagian data ke dalam sejumlah klaster berdasarkan kedekatan nilai, K-Means mampu menghasilkan pengelompokan yang jelas dan mudah dipahami. Implementasi algoritma ini semakin praktis dengan bantuan perangkat lunak RapidMiner, sebuah platform data science berbasis GUI yang memungkinkan analisis dilakukan tanpa memerlukan kemampuan pemrograman yang rumit. Selain itu, RapidMiner[10] juga menyediakan visualisasi hasil analisis yang memudahkan interpretasi, sehingga pengguna dapat segera memanfaatkan temuan untuk mendukung pengambilan keputusan bisnis.

Berbagai penelitian terdahulu telah membuktikan efektivitas metode ini. Panjaitan dkk. (2025) mengelompokkan data penjualan[11] sembako dengan algoritma K-Means dan menemukan dua klaster[12] optimal dengan nilai DBI -0,770 yang membantu strategi manajemen stok. Senada dengan itu, Faidah dan Fatah (2025) meneliti 200 transaksi penjualan di sebuah toko sembako[13] dan menemukan bahwa produk Indomie dan Tepung Segitiga merupakan produk terlaris dalam klaster penjualan tinggi. Kristanti dkk. (2024) menerapkan K-Means pada 1.140 data pelanggan dengan parameter usia, pendapatan, dan model RFM, menghasilkan empat klaster dengan nilai Silhouette 0,716 yang menunjukkan segmentasi pelanggan cukup baik[14]. Selanjutnya, penelitian Nahjan, Heryana, dan Voutama (2023) pada toko OJ Cell menemukan pola pembelian konsumen optimal pada K=3 yang mendukung strategi promosi. Terakhir, Indriani, Irawan, dan Bahtiar (2024) memanfaatkan K-Means untuk mengelompokkan stok ke kategori cepat, sedang, dan lambat laku, sehingga mempermudah manajemen inventaris. Jika dibandingkan dengan penelitian terdahulu, terdapat sejumlah persamaan dan perbedaan yang menegaskan posisi kebaruan penelitian ini. Persamaannya, penelitian Panjaitan dkk. (2025) maupun Faidah dan Fatah (2025) sama-sama menekankan pentingnya segmentasi produk untuk mendukung strategi pengadaan dan promosi, sejalan dengan fokus penelitian ini. Namun, keduanya masih terbatas pada jumlah transaksi yang relatif kecil sehingga hasil klasterisasi hanya merefleksikan kondisi toko tertentu. Penelitian Kristanti dkk. (2024) memang sama-sama menggunakan algoritma K-Means[15], tetapi fokusnya berbeda karena diarahkan pada segmentasi pelanggan[16], bukan produk. Penelitian Nahjan, Heryana, dan Voutama (2023) menunjukkan kedekatan konteks karena membahas pola pembelian konsumen, tetapi belum mengaitkannya secara langsung dengan strategi manajemen stok. Sedangkan penelitian Indriani, Irawan, dan Bahtiar (2024) lebih menekankan pada efisiensi inventaris dengan pengelompokan stok, bukan optimalisasi segmentasi penjualan produk secara menyeluruh.[17].

Tujuan utama penelitian ini adalah mengimplementasikan algoritma K-Means Clustering berbasis RapidMiner[18] untuk mengidentifikasi pola dan menganalisis segmentasi strategis dalam penjualan produk.

Melalui pendekatan ini, penelitian berupaya mengelompokkan produk berdasarkan kesamaan pola penjualan[19] sehingga dapat diperoleh informasi strategis mengenai produk dengan tingkat penjualan tinggi, sedang, maupun rendah. Segmentasi semacam ini diharapkan mampu membantu perusahaan ritel dalam menyusun kebijakan yang lebih tepat, baik terkait strategi pemasaran[20], promosi, maupun pengelolaan persediaan barang. Kontribusi yang diharapkan dari penelitian ini terbagi dalam dua aspek utama. Pertama, dari sisi akademis, penelitian ini memperkaya kajian literatur mengenai penerapan data mining di sektor ritel, khususnya melalui pemanfaatan algoritma K-Means[19] sebagai salah satu metode pengelompokan yang sederhana namun efektif. Hasil penelitian ini juga dapat menjadi referensi bagi studi lanjutan yang menyoroti analisis data penjualan menggunakan pendekatan serupa. Kedua, dari sisi praktis, penelitian ini diharapkan memberikan manfaat nyata bagi pelaku usaha ritel berupa rekomendasi strategis dalam meningkatkan efisiensi operasional. Temuan penelitian ini dapat digunakan untuk mengurangi risiko shrinkage, memperbaiki sistem manajemen stok, dan mengarahkan strategi pemasaran yang lebih tepat sasaran. Dengan demikian, penelitian ini diharapkan mampu mendukung peningkatan daya saing bisnis di tengah kompetisi pasar yang semakin ketat.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Gambar 1 berikut merupakan tahapan dari penelitian yang dilakukan.

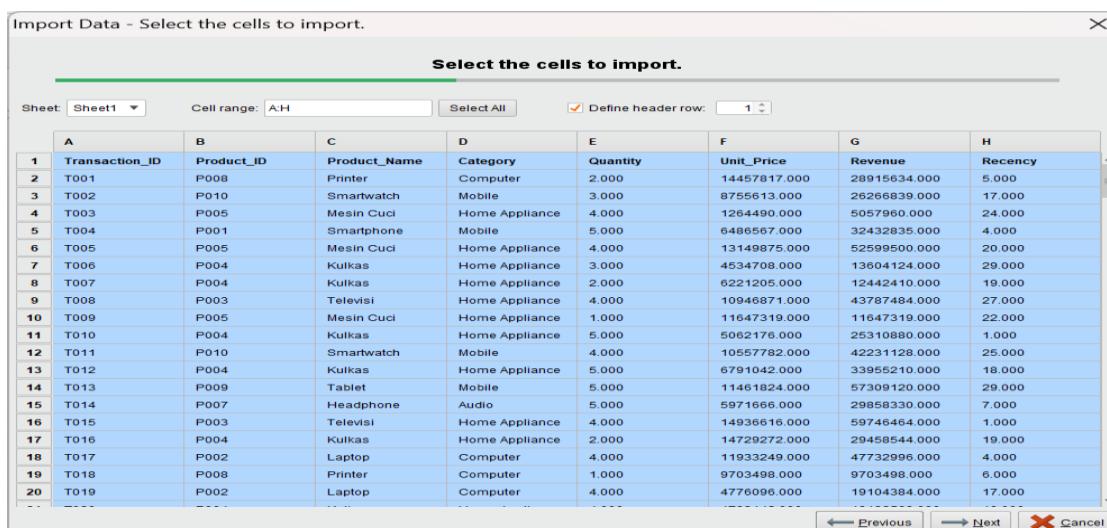


**Gambar 1.** Tahapan Penelitian

Gambar 1 menunjukkan alur metodologi penelitian, mulai dari pengumpulan dan preprocessing data, penerapan algoritma K-Means Clustering, hingga interpretasi hasil untuk menghasilkan strategi bisnis serta kesimpulan dan rekomendasi akhir.

#### 2.1.1 Pengumpulan Data

Tahap pengumpulan data dilakukan dengan mengimpor dataset penjualan produk elektronik dari sistem point-of-sale (POS) dan database transaksi Toko XYZ dalam bentuk file CSV atau Excel. Dataset ini berisi informasi penting seperti Transaction\_ID, tanggal transaksi, identitas dan kategori produk, jumlah penjualan, harga satuan, total pendapatan, serta nilai recency yang dihitung berdasarkan selisih waktu antara transaksi terakhir dengan tanggal cut-off analisis. Pada tahap ini juga dilakukan pemeriksaan awal terhadap kualitas data, seperti pengecekan duplikasi, konsistensi tipe data, nilai yang hilang, serta validitas nilai numerik untuk memastikan dataset layak digunakan pada tahap preprocessing. Seluruh data yang dikumpulkan kemudian didokumentasikan, termasuk periode pengamatan, sumber data, serta asumsi yang digunakan, sehingga proses analisis selanjutnya dapat dilakukan secara terstruktur dan dapat dipertanggungjawabkan.



1	Transaction_ID	Product_ID	Product_Name	Category	Quantity	Unit_Price	Revenue	Recency
2	T001	P008	Printer	Computer	2.000	14457817.000	28915634.000	5.000
3	T002	P010	Smartwatch	Mobile	3.000	8755613.000	26266839.000	17.000
4	T003	P005	Mesin Cuci	Home Appliance	4.000	1264490.000	5057960.000	24.000
5	T004	P001	Smartphone	Mobile	5.000	6486567.000	32432835.000	4.000
6	T005	P005	Mesin Cuci	Home Appliance	4.000	13149875.000	52599500.000	20.000
7	T006	P004	Kulkas	Home Appliance	3.000	4534708.000	13604124.000	29.000
8	T007	P004	Kulkas	Home Appliance	2.000	6221205.000	12442410.000	19.000
9	T008	P003	Televisi	Home Appliance	4.000	10946871.000	43787484.000	27.000
10	T009	P005	Mesin Cuci	Home Appliance	1.000	11647319.000	11647319.000	22.000
11	T010	P004	Kulkas	Home Appliance	5.000	5062176.000	25310880.000	1.000
12	T011	P010	Smartwatch	Mobile	4.000	10557782.000	42231128.000	25.000
13	T012	P004	Kulkas	Home Appliance	5.000	6791042.000	33955210.000	18.000
14	T013	P009	Tablet	Mobile	5.000	11461824.000	57309120.000	29.000
15	T014	P007	Headphone	Audio	5.000	5971666.000	29858330.000	7.000
16	T015	P003	Televisi	Home Appliance	4.000	14936616.000	59746464.000	1.000
17	T016	P004	Kulkas	Home Appliance	2.000	14729272.000	29458544.000	19.000
18	T017	P002	Laptop	Computer	4.000	11933249.000	47732996.000	4.000
19	T018	P008	Printer	Computer	1.000	9703498.000	9703498.000	6.000
20	T019	P002	Laptop	Computer	4.000	4776096.000	19104384.000	17.000

**Gambar 2.** Tahap Pengumpulan Data

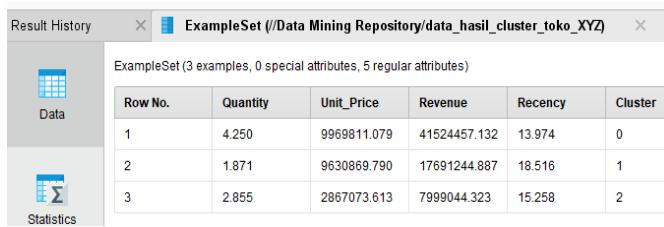
Gambar 2 merupakan tahap pengumpulan data dengan mengimpor dataset penjualan produk elektronik Toko XYZ yang berisi beberapa variabel utama, yaitu Transaction\_ID, Product\_ID, Product\_Name, Category, Quantity, Unit\_Price, Revenue, dan Recency. Variabel-variabel tersebut merepresentasikan identitas transaksi, identitas produk, kategori produk, jumlah unit terjual, harga satuan, total pendapatan per transaksi, serta jarak waktu sejak transaksi terakhir dilakukan. Data ini diperoleh dari catatan penjualan harian yang kemudian disusun

dalam format tabel untuk memudahkan proses analisis. Tahap pengumpulan data ini menjadi dasar penting dalam penelitian karena kualitas dan kelengkapan data sangat menentukan akurasi hasil analisis clustering yang akan dilakukan pada tahap berikutnya.

### 2.1.2 Preprocessing Data

Pada tahap preprocessing, data dirangkum melalui agregasi untuk memperoleh nilai inti seperti total penjualan, pendapatan, dan recency. Normalisasi menggunakan Z-Score dipilih karena lebih tahan terhadap outlier dibanding Min-Max. Data hilang ditangani dengan menghapus entri yang tidak memiliki informasi kunci dan mengimputasi nilai kosong non-kritis menggunakan median.

#### 2.1.2.1 Filter



Row No.	Quantity	Unit_Price	Revenue	Recency	Cluster
1	4.250	9969811.079	41524457.132	13.974	0
2	1.871	9630869.790	17691244.887	18.516	1
3	2.855	2867073.613	7999044.323	15.258	2

**Gambar 3.** Proses Clustering

Gambar 3 menampilkan output hasil proses clustering di RapidMiner, di mana dataset yang telah dinormalisasi dan diproses sebelumnya dikelompokkan menggunakan algoritma K-Means. Tabel menunjukkan tiga contoh data produk beserta nilai atribut Quantity, Unit\_Price, Revenue, dan Recency yang digunakan sebagai dasar pengelompokan. Kolom “Cluster” menunjukkan bahwa setiap produk telah ditempatkan ke dalam salah satu dari tiga cluster, yaitu Cluster 0, Cluster 1, dan Cluster 2, sesuai kedekatan karakteristiknya dengan centroid masing-masing. Hasil ini menegaskan bahwa algoritma K-Means berhasil memisahkan produk ke dalam segmen penjualan yang berbeda, sehingga dapat digunakan untuk analisis performa dan penyusunan strategi bisnis selanjutnya.

#### 2.1.2.2 Agregasi

Tabel berikut menampilkan data yang telah dirangkum dan dinormalisasi menggunakan Z-Score sebagai input untuk proses clustering. Setiap produk ditampilkan dengan atribut utama seperti total penjualan, harga rata-rata, total pendapatan, recency, dan jumlah transaksi, sehingga siap digunakan dalam penerapan algoritma K-Means.

**Tabel 1.** Aggregated Data (sebelum normalisasi)

Product ID	Total Quantity	Avg Price	Total Revenue	Avg Recency
P001	0,100694	-0.65	0,068056	-1.45
P002	01.22	0,097222	02.10	-0.65
P003	-0.10	00.25	-0.12	00.05
P004	-0.45	0,052083	-0.05	0,045139
P005	-0.75	00.15	-0.62	0,065972
P006	-1.10	-1.25	-1.35	-0.15
P007	00.20	-1.50	-0.85	-0.75
P008	-0.65	-0.35	-0.45	00.15
P009	0,055556	00.40	0,045139	-0.25
P010	-0.20	-0.75	-0.30	-0.05

Tabel 1 merupakan agregasi yang menampilkan ringkasan data penjualan produk elektronik Toko XYZ berdasarkan beberapa variabel penting yang diperoleh dari hasil pengolahan data transaksi. Setiap produk dirangkum dalam bentuk total kuantitas terjual, harga rata-rata, total pendapatan (total revenue), rata-rata recency, serta jumlah transaksi. Berdasarkan tabel, terlihat bahwa produk Laptop (P002) memberikan kontribusi pendapatan terbesar yaitu Rp2.418.000.000 meskipun jumlah kuantitas terjual (310 unit) lebih sedikit dibandingkan Smartphone (P001) yang berhasil terjual sebanyak 420 unit dengan total pendapatan Rp1.470.000.000. Produk lain seperti Tablet (P009) dan Televisi (P003) juga memberikan kontribusi signifikan dengan revenue di atas Rp900 juta.

Berdasarkan tabel agregasi, terlihat perbedaan yang cukup jelas antara produk unggulan dan produk dengan tingkat penjualan rendah. Produk unggulan ditunjukkan oleh Smartphone (P001), Laptop (P002), dan Tablet (P009). Ketiga produk ini memiliki kuantitas terjual tinggi, pendapatan (total revenue) yang besar, serta nilai recency yang rendah, menandakan produk tersebut sering dibeli konsumen. Misalnya, Laptop meskipun hanya terjual 310 unit, mampu menyumbang pendapatan terbesar yaitu Rp2,4 miliar berkat harga jual rata-rata yang tinggi. Smartphone menjadi produk dengan volume penjualan terbanyak (420 unit) sekaligus memberikan revenue

Rp1,47 miliar. Tablet juga memperlihatkan kinerja yang kuat dengan revenue Rp1,09 miliar. Sebaliknya, produk yang tergolong kurang laku dapat dilihat pada Mesin Cuci (P005), Speaker (P006), Headphone (P007), dan Printer (P008). Produk-produk ini memiliki kuantitas penjualan lebih rendah, rata-rata harga yang relatif kecil, serta recency lebih tinggi sehingga menunjukkan jarang dibeli. Speaker misalnya, hanya menyumbang Rp114 juta dengan 95 unit terjual, sementara Headphone meskipun terjual 210 unit hanya menghasilkan Rp178,5 juta karena harga satuannya rendah. Mesin Cuci juga termasuk dalam kategori slow-moving dengan 18,2 hari rata-rata recency, yang artinya pembeliannya jarang terjadi.

Perbandingan ini memperlihatkan bahwa produk unggulan menjadi motor penggerak utama pendapatan toko dan perlu mendapat prioritas dalam pengadaan serta strategi promosi, sedangkan produk dengan penjualan rendah memerlukan evaluasi apakah akan tetap dipertahankan atau dialihkan melalui strategi khusus seperti clearance sale atau promosi bundling.

### 2.1.2.3 Normalisasi

Tabel berikut menyajikan hasil normalisasi data menggunakan metode Z-Score, yang bertujuan menyamakan skala tiap variabel agar tidak ada atribut yang mendominasi dalam proses clustering. Nilai pada tabel menunjukkan skor standar dari total kuantitas, harga rata-rata, total pendapatan, recency, dan jumlah transaksi untuk setiap produk

**Tabel 2.** Normalized Data (setelah standardisasi Z-Score)

Product ID	Total_Quantity	Avg_Price	Total_Revenue	Avg_Recency	Transaction_Count
P001	0,100694	-0.65	0,068056	-1.45	0,091667
P002	01.22	0,097222	02.10	-0.65	01.25
P003	-0.10	00.25	-0.12	00.05	-0.05
P004	-0.45	0,052083	-0.05	0,045139	-0.35
P005	-0.75	00.15	-0.62	0,065972	-0.65
P006	-1.10	-1.25	-1.35	-0.15	-0.95
P007	00.20	-1.50	-0.85	-0.75	00.10
P008	-0.65	-0.35	-0.45	00.15	-0.55
P009	0,055556	00.40	0,045139	-0.25	00.55
P010	-0.20	-0.75	-0.30	-0.05	-0.05

Tabel 2 merupakan normalisasi yang menampilkan hasil transformasi data penjualan produk elektronik Toko XYZ menggunakan metode Z-Score Standardization. Proses normalisasi ini bertujuan untuk menyamakan skala antar variabel sehingga tidak ada atribut yang mendominasi dalam proses clustering. Dengan metode ini, setiap nilai variabel dikonversi menjadi skor standar berdasarkan rata-rata dan standar deviasi masing-masing atribut.

Dari tabel terlihat bahwa produk Laptop (P002) memiliki skor sangat tinggi pada variabel Total\_Revenue (2,10) dan Avg\_Price (1,80), menandakan bahwa produk ini mendominasi dalam hal pendapatan dan harga dibandingkan produk lain. Demikian pula Smartphone (P001) memperoleh skor tinggi pada Total\_Quantity (1,85) dan Transaction\_Count (1,72), menunjukkan intensitas penjualan dan frekuensi transaksi yang tinggi. Sebaliknya, produk Speaker (P006) dan Headphone (P007) memiliki skor negatif hampir pada semua variabel, mencerminkan kontribusi rendah baik dari sisi kuantitas, revenue, maupun harga rata-rata.

Normalisasi juga memperlihatkan bahwa variabel Avg\_Recency pada Smartphone bernilai -1,45, yang menunjukkan produk ini sering dibeli (semakin rendah recency berarti semakin cepat dibeli kembali), sedangkan Mesin Cuci (P005) memiliki skor 0,95 pada Avg\_Recency, menandakan jarang dibeli. Dengan hasil ini, dataset telah berada pada skala yang seimbang sehingga siap digunakan dalam penerapan algoritma K-Means Clustering untuk menghasilkan segmentasi produk yang lebih akurat.

### 2.1.3 Penerapan Algoritma K-Means Clustering

Pada tahap ini, dataset yang telah melalui preprocessing dianalisis menggunakan algoritma K-Means dengan jumlah cluster ditentukan sebanyak tiga berdasarkan hasil Elbow Method yang menunjukkan titik optimum pada K=3. Algoritma bekerja dengan menghitung jarak antar data menggunakan Euclidean Distance dan menempatkan setiap produk ke dalam cluster yang paling sesuai berdasarkan kesamaan karakteristik. Proses ini menghasilkan pengelompokan produk menjadi tiga segmen utama, yaitu produk dengan penjualan tinggi, menengah, dan rendah, yang selanjutnya dijadikan dasar untuk analisis lebih lanjut.



**Gambar 4.** Proses Analisis dengan RapidMiner

Gambar 4 menunjukkan proses awal implementasi analisis di RapidMiner. Dataset penjualan produk elektronik Toko XYZ dipanggil menggunakan operator Retrieve data\_penjualan\_toko\_XYZ, kemudian dilanjutkan dengan operator Filter Examples untuk menyaring data sesuai kriteria tertentu, seperti periode transaksi atau jenis produk. Tahap ini merupakan bagian dari preprocessing, yang bertujuan memastikan hanya data relevan yang digunakan pada analisis berikutnya, sehingga hasil clustering dengan algoritma K-Means dapat lebih akurat dan bermanfaat bagi pengambilan keputusan bisnis.

#### 2.1.4 Hasil Clustering

Dengan algoritma K-Means, data terbagi menjadi tiga cluster: Cluster 0 (High Sales & Revenue), Cluster 1 (Medium Sales & Revenue), dan Cluster 2 (Low Sales & Revenue), yang masing-masing menunjukkan karakteristik berbeda dalam pola penjualan produk.

#### 2.1.5 Interpretasi & Strategi Bisnis

Setiap cluster dianalisis untuk merumuskan strategi bisnis, misalnya menjaga stok dan promosi bundling untuk produk unggulan (Cluster 0), melakukan promosi musiman untuk produk sedang (Cluster 1), serta clearance sale dan evaluasi stok untuk produk slow-moving (Cluster 2).

#### 2.1.6 Kesimpulan dan Rekomendasi

Tahapan akhir penelitian memberikan kesimpulan bahwa K-Means Clustering efektif dalam memetakan pola penjualan produk dan menghasilkan rekomendasi strategis untuk mendukung efisiensi operasional dan pengambilan keputusan bisnis di Toko XYZ.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Hasil Pengujian

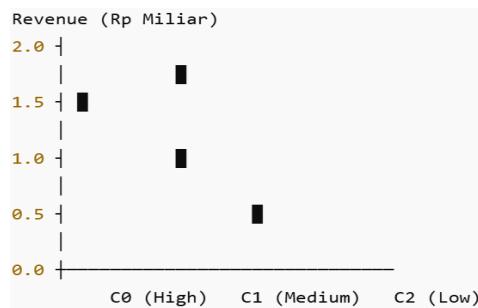
Tabel 3 berikut menyajikan nilai centroid final dari hasil proses K-Means, yang menggambarkan rata-rata karakteristik setiap cluster berdasarkan variabel penjualan yang telah dinormalisasi. Nilai-nilai ini menjadi representasi utama untuk membedakan profil Cluster High, Medium, dan Low Sales.

**Tabel 3.** Tabel Centroid Final

Cluster	Total Quantity	Avg Price	Total Revenue	Avg Recency
Cluster 0 (High Sales)	01.29	00.52	01.24	-0.78
Cluster 1 (Medium Sales)	-0.25	-0.08	-0.15	00.22
Cluster 2 (Low Sales)	-0.88	-0.54	-0.82	00.23

Tabel 3 merupakan nilai centroid final dari tiga cluster hasil metode K-Means. Cluster 0 menunjukkan nilai tertinggi pada total kuantitas, harga rata-rata, dan total pendapatan sehingga menggambarkan kelompok produk dengan performa penjualan terbaik. Cluster 1 memiliki nilai mendekati rata-rata dan mencerminkan produk dengan penjualan stabil. Sementara itu, Cluster 2 menunjukkan nilai terendah pada hampir semua atribut, menandakan produk dengan penjualan rendah dan frekuensi transaksi yang minim.

#### 3.1.1 Revenue (Rp Miliar)



**Gambar 5.** Revenue

Gambar 5 menunjukkan hasil clustering menunjukkan bahwa Cluster 0 merupakan kelompok produk paling dominan sehingga perlu menjadi prioritas dalam pengelolaan stok dan strategi promosi. Cluster 1 memiliki kinerja menengah dan membutuhkan upaya promosi tambahan agar tetap kompetitif. Sementara itu, Cluster 2 berisi produk dengan penjualan rendah yang perlu dievaluasi, bahkan sebagian mungkin memerlukan strategi clearance.

##### 1. Kontribusi Revenue Tiap Klaster

Total pendapatan yang dihasilkan dari proses clustering menunjukkan bahwa Cluster 0 memberikan kontribusi terbesar, yaitu Rp4.980.000.000, disusul Cluster 1 dengan Rp2.287.000.000. Sementara itu, Cluster 2

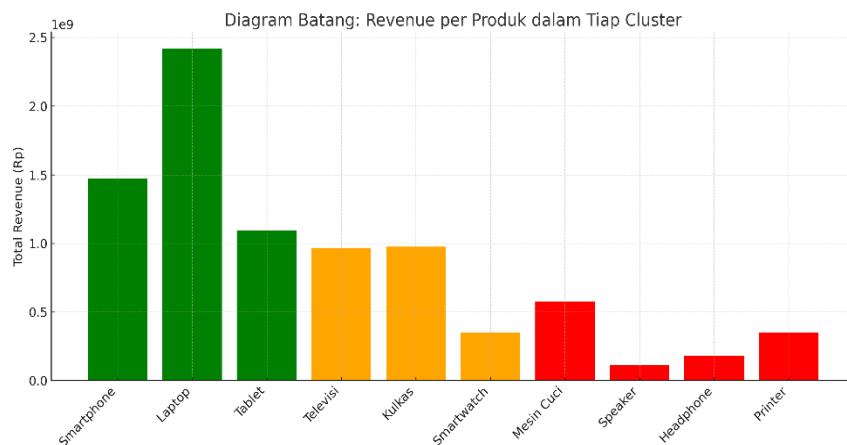
menyumbang Rp1.219.500.000, sehingga total keseluruhan revenue mencapai Rp8.486.500.000. Temuan ini menegaskan perbedaan signifikan antar klaster dalam memberikan kontribusi terhadap pendapatan toko.

## 2. Diagram Pie (Proporsi Revenue)



**Gambar 6.** Proporsi Revenue

Gambar 6 menunjukkan bahwa Cluster 0 sebagai kelompok produk dengan penjualan tinggi menyumbang hampir 60% dari total revenue, menjadikannya prioritas utama dalam strategi stok dan promosi, terutama untuk Smartphone, Laptop, dan Tablet. Cluster 1 memberikan kontribusi sekitar 27% melalui produk seperti Televisi, Kulkas, dan Smartwatch, sehingga memerlukan penguatan promosi agar tetap kompetitif. Sementara itu, Cluster 2 hanya berkontribusi 14%, mencerminkan performa rendah pada produk seperti Speaker, Headphone, Printer, dan Mesin Cuci, sehingga perlu strategi khusus seperti promosi agresif atau program clearance.



**Gambar 7.** Revenue per Produk dalam Tiap Cluster

Gambar 7 menggambarkan Cluster 0 (High Sales) yang menunjukkan bahwa Laptop dan Smartphone menjadi penyumbang revenue terbesar, disusul oleh Tablet sebagai produk pendukung. Pada Cluster 1 (Medium Sales), kontribusi Televisi dan Kulkas tampak relatif seimbang, sementara Smartwatch memberikan kontribusi yang jauh lebih kecil. Adapun Cluster 2 (Low Sales) memperlihatkan bahwa Mesin Cuci masih memberikan pendapatan lebih tinggi dibandingkan Printer, Headphone, dan Speaker yang kontribusinya sangat rendah.

Dengan grafik ini, manajemen Toko XYZ dapat langsung melihat produk mana yang dominan di tiap klaster, sehingga strategi bisa lebih terarah (misalnya fokus pada Laptop & Smartphone di cluster high, dan promo clearance untuk Speaker & Headphone di cluster low).

### 3.1.2 Hasil Clustering

Hasil penerapan algoritma K-Means Clustering terhadap data yang telah dinormalisasi menghasilkan tiga kelompok utama produk elektronik Toko XYZ.

#### a. Cluster 0 (High Sales & High Revenue)

Cluster ini berisi produk-produk unggulan dengan kuantitas terjual tinggi, pendapatan besar, frekuensi transaksi sering, serta recency rendah yang menandakan produk sering dibeli. Produk yang termasuk dalam cluster ini adalah Smartphone (P001), Laptop (P002), dan Tablet (P009). Produk-produk tersebut menjadi motor penggerak utama penjualan toko dengan kontribusi revenue hampir 60% dari total keseluruhan.

#### b. Cluster 1 (Medium Sales & Moderate Revenue)

Produk dalam cluster ini memiliki tingkat penjualan menengah dengan kontribusi pendapatan cukup stabil, meskipun tidak sebesar cluster pertama. Anggota cluster ini adalah Televisi (P003), Kulkas (P004), dan Smartwatch (P010). Karakteristiknya menunjukkan adanya potensi pasar yang cukup baik, sehingga strategi bisnis yang sesuai adalah promosi musiman, bundling dengan produk populer, serta evaluasi permintaan berdasarkan tren tertentu.

#### c. Cluster 2 (Low Sales & Low Revenue)

Cluster terakhir terdiri atas produk dengan kuantitas penjualan rendah, kontribusi revenue kecil, dan recency relatif tinggi, menandakan produk jarang dibeli. Anggota cluster ini meliputi Mesin Cuci (P005), Speaker (P006), Headphone (P007), dan Printer (P008). Produk dalam kelompok ini dikategorikan sebagai slow-moving products yang berisiko menimbulkan penumpukan stok. Strategi yang dapat diterapkan adalah program

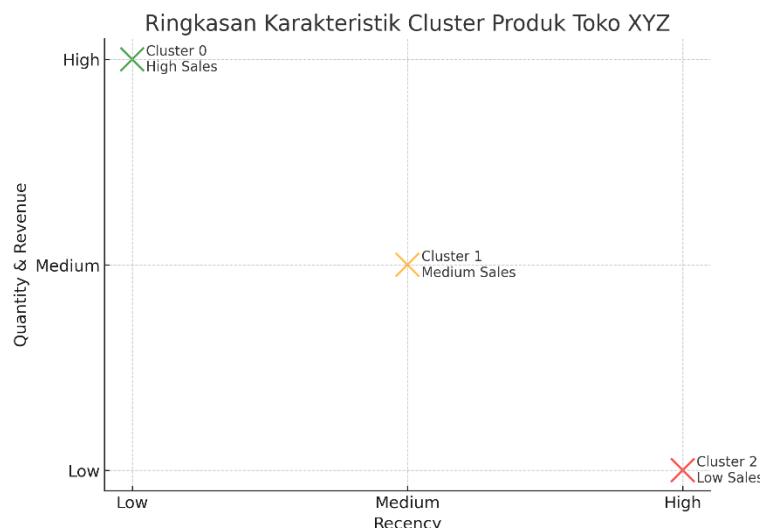
clearance sale, promosi bundling dengan produk high sales, serta evaluasi keberlanjutan produk dalam katalog toko.

Secara keseluruhan, hasil clustering ini memudahkan manajemen Toko XYZ dalam memahami perbedaan karakteristik penjualan antar produk. Segmentasi yang dihasilkan menjadi dasar pengambilan keputusan yang lebih tepat sasaran, baik dalam pengadaan stok, promosi, maupun perencanaan strategi pemasaran jangka panjang.

**Tabel 4.** Ringkasan Karakteristik Klaster

Cluster	Karakteristik Utama	Produk Dominan	Strategi Bisnis
0	Tinggi qty, revenue, transaksi; recency rendah	Smartphone, Laptop, Tablet	Prioritas stok, bundling, promo
1	Sedang qty & revenue; cukup stabil	TV, Kulkas, Smartwatch	Promo musiman, paket bundling
2	Rendah qty & revenue; recency tinggi (slow-moving)	Mesin Cuci, Speaker, Headphone, Printer	Clearance, evaluasi stok

Tabel 4 merupakan ringkasan karakteristik yang menyajikan hasil segmentasi produk elektronik Toko XYZ ke dalam tiga cluster utama berdasarkan kuantitas penjualan, pendapatan (revenue), frekuensi transaksi, dan nilai recency. Hasil segmentasi menunjukkan bahwa Cluster 0 ditandai dengan kuantitas, pendapatan, dan frekuensi transaksi yang tinggi serta recency rendah, yang berarti produk sering dibeli. Produk dominan dalam kelompok ini adalah Smartphone, Laptop, dan Tablet, sehingga strategi yang disarankan adalah menjaga ketersediaan stok (safety stock), melakukan promosi bundling, dan memperkuat strategi pemasaran agar kontribusi penjualan tetap maksimal. Sementara itu, Cluster 1 memiliki kuantitas dan revenue pada tingkat sedang dengan kondisi penjualan relatif stabil. Produk yang termasuk dalam cluster ini adalah Televisi, Kulkas, dan Smartwatch, dengan strategi bisnis yang lebih tepat berupa promosi musiman seperti diskon pada periode tertentu dan paket bundling untuk meningkatkan daya tarik penjualan. Adapun Cluster 2 terdiri atas produk dengan kuantitas dan pendapatan rendah serta recency tinggi, menandakan produk jarang dibeli atau tergolong slow-moving. Produk dalam kelompok ini adalah Mesin Cuci, Speaker, Headphone, dan Printer, yang membutuhkan strategi berupa program clearance sale, evaluasi ulang kelayakan produk, serta penggabungan dengan produk laku dalam promosi bundling. Secara keseluruhan, segmentasi ini membantu manajemen toko dalam membedakan prioritas penanganan setiap cluster, sehingga strategi pengadaan, promosi, dan pengelolaan stok dapat lebih tepat sasaran.



**Gambar 8.** Ringkasan Karakteristik Cluster

Gambar 8 merupakan grafik yang menggambarkan posisi tiga klaster produk berdasarkan hubungan antara recency dengan quantity dan revenue. Cluster 0 berada pada area recency rendah namun quantity dan revenue tinggi, menunjukkan bahwa produk dalam kelompok ini sangat laris dan sering dibeli. Cluster 1 terletak di bagian recency menengah dengan quantity dan revenue moderat, menandakan produk dengan performa cukup stabil. Sementara itu, Cluster 2 berada pada recency tinggi dan quantity serta revenue rendah, sehingga produk dalam klaster ini jarang dibeli dan memiliki kinerja penjualan paling lemah. Grafik ini mempertegas perbedaan karakteristik tiap klaster sehingga memudahkan penentuan strategi bisnis yang tepat.

### 3.2 Interpretasi dan Strategi Bisnis

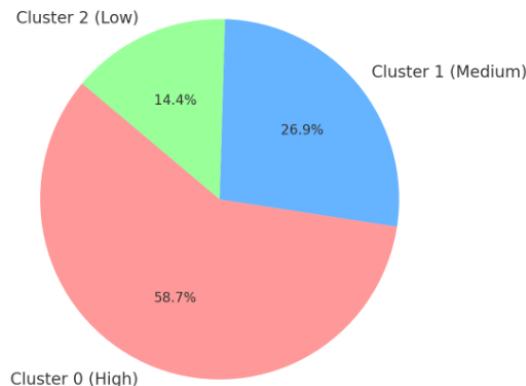
Tabel berikut menyajikan ringkasan karakteristik utama dari setiap cluster beserta strategi bisnis yang direkomendasikan. Pembagian ini membantu menggambarkan perbedaan performa penjualan antar produk, mulai dari kelompok dengan penjualan tinggi hingga produk yang tergolong lambat terjual. Melalui ringkasan ini,

strategi yang tepat dapat dirumuskan untuk masing-masing cluster agar pengelolaan stok, promosi, dan perencanaan pemasaran dapat dilakukan secara lebih efektif.

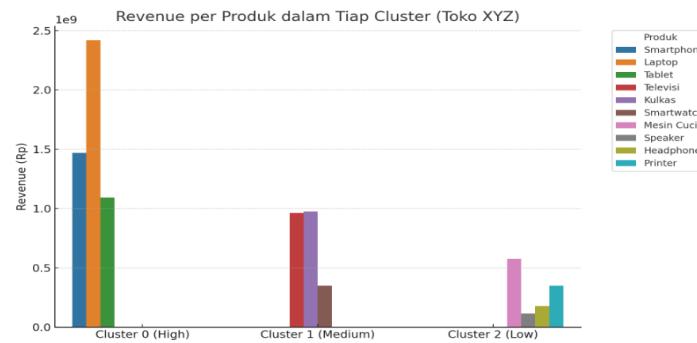
**Tabel 5.** Ringkasan Strategi Bisnis Per Produk

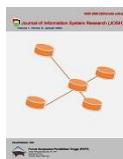
Cluster	Produk	Karakteristik	Strategi Bisnis
<b>Cluster 0 (High Sales)</b>	Smartphone, Laptop, Tablet	Penjualan tinggi, revenue besar, transaksi sering, recency rendah (sering dibeli)	<ul style="list-style-type: none"> <li>- Tingkatkan stok safety level</li> <li>- Promosi bundling (misalnya Laptop + Printer)</li> <li>- Fokus pemasaran online &amp; iklan premium</li> </ul>
<b>Cluster 1 (Medium Sales)</b>	Televisi, Kulkas, Smartwatch	Penjualan menengah, kontribusi revenue cukup signifikan, stabil namun tidak dominan	<ul style="list-style-type: none"> <li>- Promo musiman (misalnya diskon Lebaran/Natal)</li> <li>- Paket bundling (TV + Speaker)</li> <li>- Evaluasi tren musiman untuk Smartwatch</li> </ul>
<b>Cluster 2 (Low Sales)</b>	Mesin Cuci, Speaker, Headphone, Printer	Penjualan rendah, revenue kecil, recency tinggi (jarang dibeli)	<ul style="list-style-type: none"> <li>- Program clearance sale untuk stok lama</li> <li>- Bundling produk slow-moving dengan produk high sales</li> <li>- Evaluasi keberlanjutan produk di katalog</li> </ul>

Tabel 5 menggambarkan Cluster 0 yang merupakan kelompok produk andalan yang memberikan kontribusi terbesar sehingga perlu menjadi prioritas dalam pengelolaan stok dan strategi penjualan. Sementara itu, Cluster 1 terdiri atas produk dengan kinerja penjualan yang stabil dan masih memiliki potensi peningkatan melalui promosi musiman maupun program bundling. Adapun Cluster 2 berisi produk berisiko dengan penjualan rendah sehingga memerlukan langkah strategis seperti clearance atau diversifikasi untuk mencegah penumpukan stok dan mengoptimalkan ruang penyimpanan

**Kontribusi Revenue per Cluster**

**Gambar 9.** Grafik Kontribusi Revenue per Cluster

Pie chart pada Gambar 9 menunjukkan distribusi kontribusi revenue dari ketiga klaster hasil analisis K-Means. Terlihat jelas bahwa Cluster 0 (High Sales) mendominasi dengan kontribusi sebesar hampir 60% dari total revenue. Hal ini menegaskan bahwa produk dalam klaster ini, yaitu Smartphone, Laptop, dan Tablet, menjadi motor penggerak utama pendapatan Toko XYZ. Selanjutnya, Cluster 1 (Medium Sales) memberikan kontribusi sekitar 27% melalui produk seperti Televisi, Kulkas, dan Smartwatch, yang menunjukkan posisi penting meskipun tidak mendominasi. Adapun Cluster 2 (Low Sales) hanya menyumbang sekitar 14%, sehingga dapat diidentifikasi sebagai kelompok produk dengan potensi risiko slow-moving inventory.


**Gambar 10.** Grafik Revenue per Produk Tiap Cluster



Gambar 2 memperlihatkan perbandingan revenue antar produk dalam setiap klaster. Pada Cluster 0 (High Sales), produk Laptop menempati posisi tertinggi dalam menyumbang revenue, diikuti oleh Smartphone dan Tablet. Hal ini menunjukkan bahwa ketiga produk ini tidak hanya populer, tetapi juga memberikan nilai strategis yang besar bagi toko. Sementara itu, Cluster 1 (Medium Sales) didominasi oleh Televisi dan Kulkas yang memiliki kontribusi relatif seimbang, sedangkan Smartwatch masih tergolong produk dengan penjualan terbatas. Pada Cluster 2 (Low Sales), produk Mesin Cuci masih memberikan kontribusi lebih besar dibandingkan Printer, Headphone, dan Speaker, yang memiliki revenue jauh lebih kecil. Hasil ini menunjukkan bahwa strategi promosi agresif atau program clearance sangat diperlukan bagi produk-produk dalam klaster ini.

## 4. KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan algoritma K-Means Clustering berbasis RapidMiner mampu menghasilkan segmentasi produk elektronik yang lebih akurat bagi Toko XYZ. Melalui analisis terhadap data penjualan yang telah diproses dan dinormalisasi, penelitian ini berhasil mengidentifikasi tiga klaster utama, yakni produk dengan penjualan tinggi (Smartphone, Laptop, Tablet), penjualan menengah (Televisi, Kulkas, Smartwatch), serta penjualan rendah (Mesin Cuci, Speaker, Headphone, Printer). Hasil segmentasi tersebut memberikan gambaran yang jelas mengenai perbedaan karakteristik setiap kelompok produk, sehingga dapat menjadi dasar bagi manajemen dalam menyusun strategi yang lebih tepat sasaran, mulai dari prioritas pengadaan, penguatan promosi musiman dan bundling, hingga program clearance bagi produk slow-moving. Selain memberikan kontribusi akademis dalam kajian data mining pada sektor ritel, temuan ini juga memberikan manfaat praktis dalam meningkatkan efisiensi operasional serta mendukung keputusan bisnis berbasis data. Ke depan, penelitian dapat dikembangkan dengan menambah variabel seperti margin keuntungan atau perilaku pelanggan, membandingkan algoritma clustering lain, serta menggunakan dataset yang lebih besar agar segmentasi yang dihasilkan semakin komprehensif dan relevan bagi industri ritel elektronik yang terus berkembang.

## REFERENCES

- [1] I. Shantilawati, S. Zebua, and R. Tarmizi, "Penggunaan Digital Marketing Dalam Meningkatkan Penjualan Bisnis Retail," JMARI, vol. 5, no. 1, pp. 30–37, Feb. 2024, doi: 10.33050/jmari.v5i1.3130.
- [2] F. M. Alexander, A. Sasmitapura, M. Michael, and M. Meythi, "Faktor Finansial dan Nilai Perusahaan Ritel Indonesia," Jurnal Akuntansi Kontemporer, vol. 15, no. 3, pp. 155–169, 2023, doi: 10.33508/jako.v15i3.4741.
- [3] N. Qomariah, "The Role of Promotion and Service Quality in Increasing Consumer Satisfaction and Loyalty in Pawnshops," Journal of Economics, Finance And Management Studies, vol. 04, no. 10, Oct. 2021, doi: 10.47191/jefms/v4-i10-17.
- [4] M. I. Faizah and Z. Fatah, "Clustering K-Means Dengan Rapidminer Untuk Identifikasi Produk Terlaris," JUSIFOR : Jurnal Sistem Informasi dan Informatika, vol. 4, no. 1, pp. 25–33, Jun. 2025, doi: 10.70609/jusifor.v4i1.5821.
- [5] S. N. Lathifah and Z. F. Azzahra, "AI-Driven Customers Segmentation Using K-Means Clustering," G-Tech: Jurnal Teknologi Terapan, vol. 9, no. 1, pp. 320–329, Jan. 2025, doi: 10.70609/gtech.v9i1.6202.
- [6] M. Miranda and S. Srian, "Implementation of K-Means Clustering in Grouping Sales Data at Zura Mart," Journal of Applied Informatics and Computing, vol. 9, no. 2, pp. 547–555, Apr. 2025, doi: 10.30871/jaic.v9i2.9160.
- [7] L. Fernando and M. I. Fianty, "Optimizing Motorcycle Sales: Enhancing Customer Segmentation with K-Means Clustering and Data Mining Techniques," Journal of Information Systems and Informatics, vol. 6, no. 3, pp. 1484–1498, Sep. 2024, doi: 10.51519/journalisi.v6i3.799.
- [8] B. I. Nugroho, A. Rafhina, P. S. Ananda, and G. Gunawan, "Customer segmentation in sales transaction data using k-means clustering algorithm," Journal of Intelligent Decision Support System (IDSS), vol. 7, no. 2, pp. 130–136, Jun. 2024, doi: 10.35335/idss.v7i2.236.
- [9] A. Nugraha, Y. Amelia Effendi, N. Nicholas, Z. Tao, M. Afifuddin, and N. Nuzulita, "K-Means clustering interpretation using recency, frequency, and monetary factor for retail customers segmentation," TELKOMNIKA (Telecommunication Computing Electronics and Control), vol. 23, no. 2, p. 435, Apr. 2025, doi: 10.12928/telkomnika.v23i2.26044.
- [10] M. I. Faizah and Z. Fatah, "Clustering K-Means Dengan Rapidminer Untuk Identifikasi Produk Terlaris," JUSIFOR : Jurnal Sistem Informasi dan Informatika, vol. 4, no. 1, pp. 25–33, Jun. 2025, doi: 10.70609/jusifor.v4i1.5821.
- [11] M. Rafi Nahjan, Nono Heryana, and Apriade Voutama, "Implementasi Rapidminer dengan Metode Clustering K-Means untuk Analisa Penjualan Pada Toko OJ Cell," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 7, no. 1, pp. 101–104, Jan. 2023, doi: 10.36040/jati.v7i1.6094.
- [12] Mohammad Nasri Abdoel Wahid, Sudarjo, Friyato, and R. H. Tirtosetianto, "Analisis Klaster Usaha Mikro Kecil di Jawa Timur Menggunakan Metode Agglomerative Clustering dengan Software Orange Data Mining," Akademika, vol. 22, no. 1, pp. 15–21, Feb. 2024, doi: 10.51881/jak.v22i1.125.
- [13] A. Nur Rahmi and Yosaphat Ananda Mikola, "Implementasi Algoritma Apriori untuk Menentukan Pola Pembelian pada Customer (Studi Kasus : Toko Bakoel Sembako)," Information System Journal, vol. 4, no. 1, pp. 14–19, May 2021, doi: 10.24076/infosjournal.2021v4i1.561.
- [14] H. Mawarni, G. Testiana, and M. L. Dalafranka, "Implementasi Algoritma K-Means untuk Segmentasi Pelanggan Pada PT. Bintang Multi Sarana Cabang Tugumulyo," Jurnal Komputer dan Informatika, vol. 11, no. 2, pp. 227–236, Oct. 2023, doi: 10.35508/jicon.v11i2.12478.
- [15] R. Siagian, P. Sirait, and A. Halim, "The Implementation of K-Means and K-Medoids Algorithm for Customer Segmentation on E-commerce Data Transactions," SISTEMASI, vol. 11, no. 2, p. 260, May 2022, doi: 10.32520/stmsi.v11i2.1337.



- [16] J. M. John, O. Shobayo, and B. Ogunleye, “An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market,” *Analytics*, vol. 2, no. 4, pp. 809–823, Oct. 2023, doi: 10.3390/analytics2040042.
- [17] M. I. Faidah and Z. Fatah, “Clustering K-Means Dengan Rapidminer Untuk Identifikasi Produk Terlaris,” *JUSIFOR : Jurnal Sistem Informasi dan Informatika*, vol. 4, no. 1, pp. 25–33, Jun. 2025, doi: 10.70609/jusifor.v4i1.5821.
- [18] N. Dwitiyanti, Siti Ayu Kumala, and Shinta Dwi Handayani, “Comparative Study of Earthquake Clustering in Indonesia Using K-Medoids, K-Means, DBSCAN, Fuzzy C-Means and K-AP Algorithms,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 6, pp. 768–778, Dec. 2024, doi: 10.29207/resti.v8i6.5514.
- [19] A. Yahya and R. Kurniawan, “Implementasi Algoritma K-Means untuk Pengelompokan Data Penjualan Berdasarkan Pola Penjualan,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 1, pp. 350–358, Jan. 2025, doi: 10.57152/malcom.v5i1.1773.
- [20] Mifta Almaripat, Ahmad Faqih, and Ade Rizki Rinaldy, “Sales Data Classterization Analysis Using K-Means Method for Marketing Strategy Development,” *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 4, no. 2, pp. 972–976, Feb. 2025, doi: 10.59934/jaiea.v4i2.792.