

Clustering Penjualan Toko Retail Menggunakan Algoritma K-Means dalam Proses Penambangan Data

¹Eunike Charina Ibrena Tarigan, ²Angela Wita Simanullang, ³Daniel S. Simbolon, ⁴Sardo Sipayung
^{1,2,3,4}Universitas Katolik Santo Thomas
Kota Medan, Indonesia

¹*euniketrn27@gmail.com, ²angelawita38@gmail.com, ³danielsimbolon1213@gmail.com,
⁴pinarsiphom@gmail.com

*Penulis Korespondensi

Diajukan : 20/01/2026
Diterima : 30/01/2026
Dipublikasi : 30/01/2026

ABSTRAK

Perkembangan bisnis ritel yang semakin kompetitif menuntut pengelolaan data penjualan secara lebih efektif dan berbasis analisis. Data penjualan yang tersimpan dalam jumlah besar sering kali belum dimanfaatkan secara optimal untuk mendukung pengambilan keputusan strategis. Penelitian ini bertujuan untuk menerapkan teknik *data mining* dengan menggunakan algoritma K-Means dalam melakukan pengelompokan data penjualan pada toko ritel. Proses clustering dilakukan berdasarkan atribut penjualan, seperti jumlah transaksi dan volume penjualan, guna mengidentifikasi pola dan karakteristik penjualan yang serupa. Metode K-Means dipilih karena kemampuannya dalam mengelompokkan data secara efisien dan mudah diimplementasikan pada dataset berukuran besar. Berdasarkan hasil analisis, algoritma K-Means berhasil mengelompokkan produk ke dalam beberapa kluster penjualan dengan karakteristik tinggi, sedang, dan rendah. Informasi yang dihasilkan dari proses clustering ini dapat dimanfaatkan oleh pihak manajemen toko ritel sebagai dasar dalam menyusun strategi pemasaran, pengelolaan stok, serta peningkatan kinerja penjualan secara lebih terarah. Dengan demikian, penerapan algoritma K-Means dalam proses penambangan data terbukti mampu membantu pengambilan keputusan yang lebih tepat dan berbasis data pada bisnis ritel.

Kata Kunci: Data Mining, Clustering, Algoritma K-Means, Penjualan Retail, Analisis Data

I. PENDAHULUAN

Transaksi jual beli eceran atau retailing merupakan aktivitas perdagangan yang melibatkan penjual sebagai retailer dan pembeli sebagai pelanggan (Immanuel & Alfian, 2025). Retailing merupakan tahap akhir dalam rantai pasokan, di mana produk dan jasa disalurkan dari produsen atau pemasok kepada pelanggan akhir. Aktivitas retailing telah menjadi bagian yang tidak terpisahkan dari kehidupan masyarakat dan memiliki peran penting dalam perekonomian, karena menyediakan kemudahan akses bagi pelanggan dalam memenuhi kebutuhan sehari-hari terhadap berbagai produk dan jasa.

Data mining adalah proses pengolahan data secara semiotomatis yang menggunakan pendekatan statistik, matematika, kecerdasan buatan, dan machine learning untuk menemukan informasi yang bermanfaat dari kumpulan data yang besar, sehingga dapat mendukung pengambilan keputusan bisnis. Metode Metode K-Means merupakan teknik clustering non-hierarki yang mampu mengelompokkan data ke dalam beberapa kluster secara efektif (Gantara & Ali, 2023). Pada dasarnya, data mining terbagi menjadi dua jenis utama, yakni descriptive mining dan predictive mining.

Dalam data mining, clustering digunakan sebagai teknik untuk mengelompokkan data berdasarkan kesamaan karakteristik tertentu. Clustering adalah metode pengelompokan data yang mengelompokkan objek berdasarkan tingkat kemiripan yang tinggi dalam satu klaster dan perbedaan yang jelas antar klaster (Noval & Marleny, 2025). Clustering mengelompokkan data berdasarkan jarak terdekat antar objek pada suatu dataset, dengan penentuan titik awal yang dipilih secara acak (Aranta, 2021).

Metode K-Means digunakan untuk mengidentifikasi pola penjualan produk berdasarkan tingkat kemiripan data penjualan (Noval & Marleny, 2025). K-Means Clustering adalah teknik untuk mengelompokkan data ke dalam beberapa klaster dengan memperhatikan kesamaan atribut antar objek (Ika Anikah et al., 2022). Dalam konteks penjualan, metode K-Means dapat dimanfaatkan untuk mengelompokkan produk ke dalam kategori penjualan yang berbeda, seperti sangat laris, laris, dan kurang laris (Hidayati, 2023). K-Means merupakan algoritma yang banyak diteliti dan diterapkan dalam berbagai disiplin ilmu untuk keperluan analisis data (Al Rasyid et al., 2022). Penelitian ini menerapkan algoritma K-Means pada bidang penjualan atau bisnis, dengan pertimbangan kemudahan implementasi, kecepatan proses, serta fleksibilitas dalam penerapannya.

II. STUDI LITERATUR

Penelitian Terdahulu

Dalam penelitiannya, Muhammad Fikry Adiansyah (2020) membangun aplikasi berbasis Laravel untuk pengelolaan persediaan barang dengan menggunakan Davies–Bouldin Index sebagai metode validasi cluster. Hasil analisis menunjukkan bahwa dua cluster merupakan konfigurasi yang paling optimal (Adiansyah, 2020). Berbagai penelitian telah menerapkan algoritma K-Means untuk beragam keperluan analisis data, antara lain pengelompokan data penjualan, penentuan penerima beasiswa, analisis potensi wilayah, serta manajemen persediaan barang (Agustin, 2025).

Sebagai contoh, penelitian yang dilakukan oleh Aditya Agusti Achray (2020) menerapkan algoritma K-Means untuk mengelompokkan data penjualan mobil dengan tujuan merumuskan strategi promosi yang lebih efektif pada PT Honda Arista Mangga Dua (Achray, 2020). Pengembangan sistem dilakukan berbasis web menggunakan PHP dan MySQL, dengan penerapan metode SDLC dan KDD secara bertahap. Dalam penelitian oleh Ajie Prasetya dkk. (2023), metode Elbow diterapkan untuk menentukan jumlah cluster terbaik pada analisis penjualan produk, dan hasilnya menunjukkan performa clustering yang memadai (Prasetya et al., 2023).

Berdasarkan telaah terhadap penelitian-penelitian terdahulu, algoritma K-Means banyak digunakan dalam proses penambangan data agar mencapai keperluan pengelompokan data, diantaranya yaitu analisis penjualan, manajemen persediaan, hingga perumusan strategi pemasaran. Beberapa penelitian menunjukkan bahwa K-Means mampu memberikan hasil segmentasi data yang efektif berdasarkan tingkat kemiripan karakteristik yang dimiliki. Akan tetapi, beberapa penelitian tersebut masih diimplementasikan pada objek, jenis data, serta studi yang berbeda dan penelitian tersebut juga belum secara khusus membahas analisis penjualan pada toko ritel dengan detail.

Selain itu, pemanfaatan proses penambangan data secara terstruktur dalam mengidentifikasi pola penjualan, dan pengelompokan data penjualan toko ritel masih terdapat keterbatasan sehingga masih banyak riset-riset yang belum memanfaatkan hasil clustering yang optimal sebagai dasar pendukung pengambilan keputusan bisnis dalam lingkungan ritel. Oleh karena itu, penelitian ini dilakukan untuk mengisi celah penelitian tersebut dengan menerapkan algoritma K-Means dalam proses penambangan data dengan tujuan mengelompokkan data penjualan toko ritel sehingga diperoleh informasi yang lebih sistematis dan dapat membantu pihak manajemen dalam pengambilan keputusan yang lebih tepat.

III. METODE

III.1. Desain Penelitian

Dalam penelitian ini, teknik clustering pada data mining diterapkan untuk mengelompokkan data penjualan ritel berdasarkan kemiripan karakteristik penjualan.

K-Means digunakan dalam penelitian ini untuk melakukan pengelompokan data secara non-hierarki dengan mengacu pada jarak antara data dan centroid.

III.2. Objek dan Lokasi Penelitian

Objek dalam penelitian ini berupa data penjualan produk pada toko ritel. Penelitian difokuskan pada analisis pola penjualan produk berdasarkan data transaksi penjualan yang tersimpan dalam sistem pencatatan penjualan toko ritel.

III.3. Jenis dan Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari sumber tidak langsung, seperti arsip atau basis data yang telah ada.

Tabel 3.1 Jenis dan Sumber Data

No	Jenis Data	Sumber Data	Keterangan
1	Data Penjualan	Sistem transaksi toko ritel	Digunakan sebagai data utama clustering
2	Literatur	Jurnal, buku, penelitian terdahulu	Digunakan sebagai landasan teori

III.4. Atribut Data Penelitian

Atribut data yang digunakan dalam proses clustering dipilih berdasarkan relevansi terhadap tujuan penelitian, yaitu untuk mengelompokkan produk berdasarkan tingkat penjualannya.

Tabel 3.2 Atribut Data Penjualan

No	Nama Atribut	Deskripsi
1	Kode Produk	Identitas unik produk
2	Jumlah Terjual	Total unit produk yang terjual
3	Total Penjualan	Nilai total penjualan produk
4	Frekuensi Transaksi	Jumlah transaksi produk

Atribut-atribut tersebut digunakan sebagai variabel dalam proses perhitungan jarak pada algoritma K-Means.

III.5. Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Studi Dokumentasi, yaitu pengumpulan data penjualan dari arsip transaksi toko ritel.
2. Studi Literatur, yaitu pengumpulan referensi yang berkaitan dengan data mining, clustering, dan algoritma K-Means.

III.6. Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang disusun secara sistematis agar proses clustering dapat berjalan dengan baik.

a. Pengumpulan Data

Data penjualan dikumpulkan dari sistem transaksi toko ritel sesuai dengan periode yang telah ditentukan.

b. Pra-pemrosesan Data

Tahap pra-pemrosesan bertujuan untuk meningkatkan kualitas data sebelum dianalisis.

Tabel 3.3 Tahapan Pra-pemrosesan Data

No	Tahapan	Keterangan
1	Data Cleaning	Menghapus data kosong dan data ganda
2	Transformasi Data	Mengubah data ke bentuk numerik
3	Normalisasi Data	Menyamakan skala data

III.7. Penentuan Jumlah Kluster

Jumlah kluster (K) ditentukan berdasarkan tujuan penelitian, yaitu untuk mengelompokkan data penjualan ke dalam kategori penjualan tinggi, sedang, dan rendah. Penentuan jumlah kluster ini bertujuan untuk mempermudah interpretasi hasil clustering dan mendukung pengambilan keputusan bisnis.

III.8. Proses Clustering Menggunakan Algoritma K-Means

Proses clustering menggunakan algoritma K-Means dilakukan melalui beberapa tahapan, yaitu:

1. Menetapkan jumlah kluster (K) yang akan dibentuk.
2. Menentukan centroid awal secara acak.
3. Menghitung jarak antara setiap data dan centroid menggunakan metode *Euclidean Distance*.
4. Mengelompokkan data ke dalam kluster dengan jarak terdekat terhadap centroid.
5. Memperbarui nilai centroid berdasarkan rata-rata data pada setiap kluster.
6. Mengulangi seluruh proses hingga pembentukan kluster mencapai kondisi stabil.

Tabel 3.4 Tahapan Algoritma K-Means

Tahap	Deskripsi
Inisialisasi	Penentuan centroid awal
Perhitungan Jarak	Menghitung jarak data ke centroid
Pengelompokan	Menentukan kluster data
Update Centroid	Menghitung ulang pusat kluster
Konvergensi	Menghentikan iterasi

III.9. Metode Analisis Data

Metode analisis data yang digunakan adalah analisis clustering K-Means, yaitu dengan mengamati hasil pengelompokan data ke dalam masing-masing kluster. Analisis dilakukan dengan membandingkan karakteristik setiap kluster untuk mengetahui perbedaan pola penjualan antar kluster.

III.10. Alur Penelitian

Secara umum, alur penelitian ini terdiri dari:

1. Pengumpulan data penjualan
2. Pra-pemrosesan data
3. Penentuan jumlah kluster
4. Proses clustering menggunakan algoritma K-Means
5. Analisis hasil clustering
6. Penarikan kesimpulan

IV. HASIL DAN PEMBAHASAN

Deskripsi Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data penjualan toko ritel yang diperoleh dari sistem pencatatan transaksi penjualan. Data tersebut berisi informasi penjualan produk dalam periode tertentu dan telah melalui tahap seleksi sehingga hanya atribut yang relevan yang digunakan dalam proses analisis.

Sebelum dilakukan proses clustering, data penjualan terlebih dahulu dipersiapkan agar sesuai dengan kebutuhan algoritma K-Means. Data yang digunakan bersifat numerik sehingga dapat diproses secara matematis untuk menghitung jarak antar data.

Hasil Pra-Pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk meningkatkan kualitas data sebelum dianalisis lebih lanjut. Pada tahap ini dilakukan beberapa proses, antara lain pembersihan data dari nilai kosong (*missing value*), penghapusan data ganda, serta transformasi data ke dalam bentuk numerik.

Selain itu, dilakukan proses normalisasi data agar setiap atribut memiliki skala yang seimbang. Proses ini bertujuan untuk mencegah atribut dengan nilai yang lebih besar mendominasi perhitungan jarak pada algoritma K-Means. Setelah tahap pra-pemrosesan selesai, data dinyatakan siap untuk digunakan dalam proses clustering.

Penentuan Jumlah Kluster

Penentuan jumlah kluster (K) dilakukan berdasarkan tujuan penelitian, yaitu untuk mengelompokkan data penjualan produk ke dalam beberapa kategori berdasarkan tingkat penjualannya. Dalam penelitian ini, jumlah kluster ditentukan sebanyak K kluster, yang

merepresentasikan kategori penjualan tertentu, seperti penjualan tinggi, penjualan sedang, dan penjualan rendah.

Pemilihan jumlah kluster ini bertujuan untuk memudahkan analisis pola penjualan dan membantu pihak manajemen dalam memahami karakteristik setiap kelompok produk.

Hasil Proses Clustering Menggunakan Algoritma K-Means

Proses clustering dilakukan dengan menerapkan algoritma K-Means pada data penjualan yang telah dipra-pemrosesan. Pada tahap awal, centroid awal ditentukan secara acak sesuai dengan jumlah kluster yang telah ditetapkan. Selanjutnya, jarak antara setiap data penjualan dengan centroid dihitung menggunakan metode Euclidean Distance.

Berdasarkan hasil perhitungan jarak tersebut, setiap data penjualan dikelompokkan ke dalam kluster dengan jarak terdekat. Proses ini dilakukan secara berulang dengan memperbarui nilai centroid hingga tidak terjadi perubahan kluster yang signifikan. Hal ini menunjukkan bahwa proses clustering telah mencapai kondisi konvergen. Hasil clustering menunjukkan bahwa data penjualan berhasil dikelompokkan ke dalam beberapa kluster dengan karakteristik yang berbeda-beda.

Analisis

Hasil clustering yang diperoleh menunjukkan adanya perbedaan pola penjualan pada setiap kluster. Kluster dengan nilai penjualan tertinggi menunjukkan produk-produk yang memiliki tingkat penjualan tinggi dan memberikan kontribusi besar terhadap pendapatan toko. Kluster ini dapat dikategorikan sebagai produk yang paling diminati oleh pelanggan.

Kluster dengan tingkat penjualan sedang menunjukkan produk-produk yang memiliki tingkat penjualan stabil, namun tidak terlalu dominan. Produk pada kluster ini tetap memiliki peran penting dalam menjaga keseimbangan penjualan toko.

Sementara itu, kluster dengan tingkat penjualan rendah menunjukkan produk-produk yang memiliki frekuensi penjualan rendah. Informasi ini dapat digunakan oleh pihak manajemen untuk mengevaluasi strategi pemasaran, pengelolaan stok, atau bahkan mempertimbangkan pengurangan persediaan produk tertentu.

Pembahasan

Penerapan algoritma K-Means dalam penelitian ini terbukti mampu mengelompokkan data penjualan toko ritel secara efektif berdasarkan karakteristik penjualannya. Hasil pengelompokan memberikan gambaran yang lebih terstruktur mengenai pola penjualan produk, sehingga memudahkan pihak manajemen dalam memahami kondisi penjualan secara keseluruhan.

Penelitian ini memperkuat hasil studi sebelumnya yang membuktikan efektivitas algoritma K-Means dalam pengelompokan produk berbasis data penjualan. Dengan adanya pengelompokan ini, pengambilan keputusan terkait strategi pemasaran, pengelolaan stok, dan perencanaan penjualan dapat dilakukan secara lebih tepat dan berbasis data.

Selain itu, hasil clustering dapat dijadikan sebagai dasar dalam pengembangan sistem pendukung keputusan di bidang penjualan ritel, sehingga pemanfaatan data penjualan dapat dilakukan secara lebih optimal.

Implikasi Hasil Penelitian

Implikasi dari hasil penelitian ini adalah tersedianya informasi yang dapat membantu pihak manajemen toko ritel dalam:

1. Menentukan prioritas produk berdasarkan tingkat penjualan
2. Mengoptimalkan pengelolaan stok barang
3. Menyusun strategi pemasaran yang lebih tepat sasaran
4. Mendukung pengambilan keputusan bisnis berbasis data

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma K-Means dalam proses penambangan data efektif untuk mengelompokkan data penjualan produk pada toko ritel berdasarkan kemiripan karakteristik penjualannya. Proses clustering

menghasilkan beberapa kelompok penjualan yang menunjukkan perbedaan pola penjualan, seperti tingkat penjualan tinggi, sedang, dan rendah, sehingga memberikan gambaran yang lebih jelas mengenai karakteristik setiap kelompok produk.

Hasil clustering yang diperoleh dapat dimanfaatkan sebagai dasar pendukung pengambilan keputusan oleh pihak manajemen toko ritel, khususnya dalam menentukan prioritas produk, pengelolaan stok, serta penyusunan strategi pemasaran yang lebih tepat sasaran. Namun demikian, penelitian ini memiliki keterbatasan, antara lain data yang digunakan hanya berasal dari satu toko ritel dan variabel yang dianalisis masih terbatas pada data penjualan tertentu. Oleh karena itu, hasil penelitian ini belum dapat digeneralisasikan secara luas dan masih memerlukan pengembangan lebih lanjut, seperti penambahan variabel, periode data yang lebih panjang, serta penerapan metode evaluasi clustering lainnya agar diperoleh hasil yang lebih optimal dan komprehensif.

VI. REFERENSI

- Achray, A. A. (2020). *Implementasi Algoritma K-Means Untuk Mengelompokkan Data Penjualan Mobil Di Pt . Honda Arista Mangga Dua Implementasi Algoritma K-Means Untuk Mengelompokkan Data Penjualan Mobil Di Pt .*
- Adiansyah, M. F. (2020). *Aplikasi Algoritma K-Means Untuk Menentukan Persediaan Barang Berbasis Web Dengan Framework Laravel (Vol. 2).*
- Agustin. (2025). Implementasi K-Means Untuk Pengelompokan Kategori Penjualan Barang Berbasis Web. *Journal of Informatics Management and Information Technology*, 5(3), 167–176. <https://doi.org/10.47065/jimat.v5i3.610>
- Al Rasyid, H., Firmansyah, B., Soebari, K., Satria, D., & Kartika, Y. (2022). Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penjualan Produk Pada Online Shop Toko Gizi Implementation of K-Means Clustering Algorithm for Grouping Product Sales in Online Shop Toko Gizi. *Prosiding Seminar Nasional Teknologi Dan Sistem Informasi (SITASI), September*, 10–11. <http://sitasi.upnjatim.ac.id/%7C242>
- Aranta, A. (2021). *Analisis Pemilihan Cluster Optimal Dalam Segmentasi*. 18(2), 152–163.
- Gantara, N. P., & Ali, I. (2023). Penerapan Metode K-Means Clustering Pada Penjualan Barang Di Sports Station. *E-Link: Jurnal Teknik Elektro Dan Informatika*, 18(1), 28. <https://doi.org/10.30587/e-link.v18i1.5339>
- Hidayati, N. (2023). *Penerapan Data Mining Untuk Clustering Pada Toko Laura Grosir Dan Eceran Abstrak PENDAHULUAN Peranan usaha kecil dan menengah (UKM) dalam bidang usaha ecer dan grosir mempunyai peranan penting yang sangat besar dalam memenuhi kebutuhan pokok sehari-hari*. 2(3), 51–60.
- Ika Anikah, Agus Surip, Nela Puji Rahayu, Muhammad Harun Al- Musa, & Edi Tohidi. (2022). Pengelompokan Data Barang Dengan Menggunakan Metode K-Means Untuk Menentukan Stok Persediaan Barang. *KOPERTIP : Jurnal Ilmiah Manajemen Informatika Dan Komputer*, 4(2), 58–64. <https://doi.org/10.32485/kopertip.v4i2.120>
- Imanuel, D. A., & Alfian, G. (2025). Visualisasi Segmentasi Pelanggan Berdasarkan Atribut RFM Menggunakan Algoritma K-Means Untuk Memahami Karakteristik Pelanggan pada Toko Retail Online. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 12(2), 283–292. <https://doi.org/10.25126/jtiik.2025128619>
- Noval, M., & Marleny, F. D. (2025). *Implementasi Algoritma K-Means Untuk Analisis Pola Penjualan Pada Toko Monisa JURNAL MEDIA INFORMATIKA [JUMIN]*. 6(3), 1996–2002.
- Prasetya, A., Salkiawati, R., & D. Alexander, A. (2023). Analisis Cluster K-Means dengan Metode

Elbow untuk Menentukan Pola Penjualan Produk Traffic Room Summarecon Mal Bekasi.
Journal of Students' Research in Computer Science, 4(1), 105–118.
<https://doi.org/10.31599/jsrscs.v4i1.2480>