

Perbandingan Algoritma *K-Means* dan DBSCAN dalam Pengelompokan Data Penjualan Elektronik

DOI: <http://dx.doi.org/10.35889/progresif.v21i2.2775>

Creative Commons License 4.0 (CC BY –NC)



Tria Setyani^{1*}, Yulia Indriani², Muhammad Fadli³, Erliyan Redy Susanto⁴

^{1,2,4}Magister Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

³Ekonomi dan Bisnis, Politeknik Negeri Lampung, Bandar Lampung, Indonesia

*e-mail Corresponding: tria_setyani@teknokrat.ac.id

Abstract

This study aims to compare the performance of the K-Means and DBSCAN algorithms in clustering electronic sales data at CV Rey Gasendra. Clustering methods are used to identify sales patterns based on the transaction intensity of each product. The K-Means algorithm was implemented by determining the optimal number of clusters using the Elbow Method and Silhouette Score evaluation. Meanwhile, DBSCAN was applied using a density-based approach, which does not require specifying the number of clusters in advance and relies on eps and minPts parameters. The results showed that K-Means produced more stable and interpretable clusters with a Silhouette Score of 0.961, whereas DBSCAN excelled in detecting outliers and handling irregular data distributions, despite generating a large amount of noise. Performance evaluation using metrics such as Silhouette Score, Davies-Bouldin Index (DBI), and Sum of Squared Errors (SSE) indicated that K-Means is more suitable for structured product segmentation, while DBSCAN is more effective for exploring unusual sales patterns. These findings can support data-driven business decisions, such as marketing strategies, inventory management, and product recommendations.

Keywords: *K-Means; DBSCAN; Clustering; Sales data; Product segmentation*

Abstrak

Penelitian ini bertujuan untuk membandingkan kinerja algoritma *K-Means* dan DBSCAN dalam pengelompokan data penjualan elektronik pada CV Rey Gasendra. Metode *clustering* digunakan untuk mengidentifikasi pola penjualan berdasarkan intensitas transaksi setiap produk. Algoritma *K-Means* diimplementasikan dengan penentuan jumlah kluster optimal menggunakan metode *Elbow* dan evaluasi *Silhouette Score*. Sementara itu, DBSCAN diimplementasikan dengan pendekatan berbasis kepadatan, tanpa memerlukan jumlah kluster awal, dan menggunakan parameter eps serta *minPts*. Hasil penelitian menunjukkan bahwa *K-Means* memberikan hasil *clustering* yang lebih stabil dan mudah diinterpretasikan dengan *Silhouette Score* mencapai 0,961, sedangkan DBSCAN unggul dalam mendeteksi *outlier* dan menangani data dengan distribusi tidak beraturan, meskipun menghasilkan banyak noise. Evaluasi performa berdasarkan metrik seperti *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan *Sum of Squared Errors* (SSE) menunjukkan bahwa *K-Means* lebih cocok digunakan untuk segmentasi produk penjualan yang terstruktur, sedangkan DBSCAN lebih efektif untuk eksplorasi pola penjualan yang tidak lazim. Temuan ini dapat membantu pengambilan keputusan bisnis berbasis data, seperti strategi pemasaran, manajemen stok, dan rekomendasi produk.

Kata kunci: *K-Means; DBSCAN; Clustering; Data penjualan; Segmentasi produk*

1. Pendahuluan

Di era digital saat ini, informasi telah menjadi salah satu aset terpenting dalam pengambilan keputusan. Kecepatan dan ketepatan dalam mengolah informasi berperan besar dalam meningkatkan efisiensi dan daya saing suatu entitas, termasuk dalam bidang penjualan. Namun, banyaknya data yang tersedia sering kali belum memberikan manfaat optimal karena informasi yang relevan masih tersembunyi dalam kumpulan data yang besar dan kompleks. Oleh karena itu, pemanfaatan metode analitik, khususnya teknik data mining, menjadi semakin penting untuk

membantu proses pengambilan keputusan yang berbasis data.

CV Rey Gasendra merupakan sebuah usaha yang bergerak di bidang penjualan handphone dan berbagai perangkat elektronik lainnya seperti kulkas, televisi, dan mesin cuci. Dalam operasionalnya, perusahaan ini menghadapi permasalahan dalam mengelompokkan data penjualan produk berdasarkan pola tertentu. Masalah ini menyebabkan kesulitan dalam memahami preferensi pelanggan dan tren pasar, sehingga berpengaruh terhadap efektivitas strategi penjualan. Berdasarkan hasil analisis awal, data penjualan belum dikelompokkan secara sistematis, dan belum ada metode yang digunakan untuk mengetahui segmen pasar secara menyeluruh.

Sebagai solusi untuk mengatasi permasalahan tersebut, penerapan teknik *clustering* dalam data mining dapat digunakan untuk mengelompokkan data penjualan berdasarkan kemiripan karakteristik. Dua metode *clustering* yang umum digunakan adalah *K-Means* dan DBSCAN [1]. *K-Means* efektif dalam mengelompokkan data yang memiliki bentuk *cluster* yang jelas dan ukuran seragam, namun sensitif terhadap *outlier* dan membutuhkan jumlah *cluster* yang ditentukan di awal [2][3]. Sebaliknya, DBSCAN dapat mendeteksi *cluster* dengan bentuk *arbitrer* dan secara otomatis mengabaikan *outlier*, tanpa harus menentukan jumlah *cluster* terlebih dahulu [4][5][6]. Studi sebelumnya menunjukkan bahwa pemilihan metode *clustering* yang tepat sangat bergantung pada karakteristik data yang digunakan [7], sehingga perlu dilakukan evaluasi terhadap masing-masing metode dalam konteks data penjualan elektronik di CV Rey Gasendra.

Penelitian ini bertujuan untuk membandingkan performa algoritma *K-Means* dan DBSCAN dalam mengelompokkan data penjualan di CV Rey Gasendra. Hasil dari penelitian ini diharapkan dapat memberikan masukan dalam memilih metode *clustering* yang paling sesuai, sehingga dapat digunakan untuk mendukung strategi pemasaran, memahami perilaku konsumen, serta meningkatkan efektivitas pengelolaan data penjualan secara menyeluruh.

2. Tinjauan Pustaka

Beberapa penelitian terdahulu telah mengkaji penggunaan algoritma *clustering* untuk pengelompokan data penjualan. Algoritma *K-Means* pada data penjualan produk elektronik di Batam IT Mart dengan parameter jenis produk dan volume penjualan. Hasil penelitian tersebut menghasilkan tiga klaster, yaitu produk dengan penjualan tinggi, sedang, dan rendah, yang dapat dimanfaatkan untuk strategi promosi [8]. Penelitian lain yang membandingkan tiga metode *clustering*, yaitu *K-Means*, DBSCAN, dan Hierarchical *Clustering* pada data penjualan *handphone* di *marketplace* elektronik. Penelitian ini menunjukkan bahwa metode Hierarchical menghasilkan nilai *silhouette* terbaik sebesar 0.944, namun DBSCAN lebih efektif dalam menangani data dengan noise [9]. Memanfaatkan DBSCAN dengan bantuan metode *Elbow* untuk mengelompokkan data persediaan barang ritel, serta menilai keakuratan *cluster* berdasarkan parameter epsilon dan *minPts* [10]. Sementara itu, penelitian lain yang menerapkan *K-Means* untuk mengelompokkan data penjualan *smartphone* berdasarkan unit terjual, kapasitas memori, dan warna produk, dengan hasil evaluasi menggunakan *silhouette coefficient* mencapai 0.78 [11].

Penelitian ini memiliki perbedaan signifikan dibandingkan penelitian sebelumnya. Selain membandingkan secara langsung performa algoritma *K-Means* dan DBSCAN, penelitian ini dilakukan pada data penjualan riil CV Rey Gasendra yang memiliki karakteristik data variatif dan berpotensi mengandung *outlier*. Penelitian ini juga menggunakan parameter yang lebih kompleks, seperti frekuensi pembelian, kategori produk, serta jumlah transaksi, dan mengevaluasi hasil klaster tidak hanya dari sisi teknis (*Silhouette Score*), tetapi juga dari kesesuaiannya terhadap kebutuhan strategi bisnis perusahaan. Oleh karena itu, penelitian ini menawarkan pendekatan yang lebih aplikatif dan relevan dalam konteks pengambilan keputusan di sektor ritel elektronik lokal.

3.1 Metodologi

3.1 Algoritma Kmeans dan DBSCAN

Algoritma *K-Means* merupakan metode klasterisasi berbasis partisi yang membagi data ke dalam k klaster berdasarkan kedekatan jarak terhadap pusat klaster (*centroid*) [12]. Proses iteratif algoritma ini dimulai dengan inisialisasi jumlah klaster (k) dan penempatan *centroid* awal secara acak.

K-Means adalah salah satu metode *clustering* non-hierarki (*partitioning*) yang bertujuan untuk membagi kumpulan data ke dalam dua atau lebih kelompok [13]. Tujuan dari metode ini adalah mengelompokkan data yang memiliki karakteristik serupa ke dalam satu klaster, sementara data dengan ciri yang berbeda dimasukkan ke dalam klaster lain [14].

Secara umum, proses pengelompokan menggunakan algoritma *K-Means* dilakukan melalui langkah-langkah berikut:

1. Menentukan jumlah kluster (k) yang diinginkan.
2. Membagi data ke dalam k kluster awal.
3. Menghitung pusat kluster (*centroid*) untuk setiap kelompok berdasarkan data yang telah dikelompokkan, yang dapat dinyatakan dalam bentuk persamaan.

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j \quad (1)$$

Keterangan:

C adalah *centroid*, M adalah jumlah data, dan i menyatakan jumlah kelompok.

4. Setiap data akan dialokasikan ke *centroid* terdekat. Perhitungan jarak antar data dan *centroid* dilakukan menggunakan rumus jarak *Euclidean* sebagaimana dinyatakan dalam persamaan.

$$D(x_i, C_i) = \sqrt{\sum_{j=1}^q (x_{ij} - c_{ij})^2} \quad (2)$$

5. Kembali ke langkah 3 apabila masih ada data yang berpindah kelompok

DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) adalah salah satu algoritma awal yang memperkenalkan pendekatan *clustering* berbasis kepadatan, yang dikenal sebagai *density-based clustering* dalam ranah data mining. Prinsip kerja DBSCAN didasarkan pada identifikasi area dengan kepadatan tinggi yang saling terhubung (*density-connected*) [15]. Dalam metode ini, wilayah tertentu akan dibentuk menjadi sebuah kluster jika memenuhi jumlah minimum objek yang telah ditentukan. Sementara itu, data yang tidak termasuk dalam kluster mana pun dianggap sebagai *noise* atau data luar yang tidak sesuai dengan pola kluster yang terbentuk [16].

Langkah-langkah dalam algoritma DBSCAN dapat dijelaskan sebagai berikut:

1. Menetapkan terlebih dahulu nilai parameter *minPts* dan *eps*.
2. Memilih titik awal p secara acak dari kumpulan data.
3. Mengulangi proses untuk setiap titik dalam dataset.
4. Menghitung nilai *eps*, yaitu jarak antara titik-titik yang berada dalam jangkauan kepadatan (*density reachable*) dari titik p , dengan menggunakan rumus:

$$E(x, y) = \sqrt{\sum_{i=0}^n (X_i - Y_i)^2} \quad (3)$$

5. Jika jumlah titik dalam radius *eps* melebihi atau sama dengan *minPts*, maka titik p diklasifikasikan sebagai *core point* dan akan menjadi pusat dari kluster yang terbentuk.
6. Namun, apabila titik p hanya memiliki sedikit titik dalam jangkauannya dan tidak memenuhi syarat sebagai *core point*, serta tidak terdapat titik yang *density reachable*, maka titik tersebut dianggap sebagai *border point* atau *noise*, dan algoritma akan melanjutkan perhitungan pada titik lainnya.

3.2 Data dan Parameter

Data yang digunakan dalam penelitian ini merupakan data transaksi penjualan elektronik di CV Rey Gasendra selama periode September hingga Desember 2024. Dataset mencakup atribut seperti tanggal transaksi, jenis produk, jumlah barang terjual, harga satuan, dan total penjualan. Data yang bersifat kategorikal seperti jenis produk dikonversi ke bentuk numerik, sedangkan variabel numerik dinormalisasi menggunakan *Min-Max Scaler*. Parameter input untuk proses klusterisasi meliputi jumlah produk terjual, harga satuan, dan total penjualan. Sementara itu, output yang menjadi luaran adalah label kluster hasil pengelompokan data. Jumlah data yang digunakan sebanyak 240 entri transaksi. Data yang memiliki nilai kosong (*missing value*) atau duplikat dibersihkan pada tahap awal pra-pemrosesan.

4. Hasil dan Pembahasan

4.1 Sampel Data Penelitian

Data penelitian ini diambil dari transaksi penjualan berbagai produk elektronik di CV Rey Gasendra selama periode September hingga Desember 2024. Dataset terdiri dari total 240 entri transaksi dan mencakup lebih dari 60 jenis produk, termasuk *handphone*, televisi, mesin cuci, kulkas, dan perangkat elektronik lainnya. Setiap kolom mewakili satu jenis produk dengan nilai berupa jumlah unit terjual. Data awal telah melalui proses pembersihan, seperti penghapusan nilai

kosong, baris duplikat, serta data tidak relevan, guna memastikan kualitas dan konsistensi data sebelum dilakukan proses analisis klusterisasi.

Table 1 Data Sampel Penelitian

No	Produk	Jumlah Terjual	Harga Satuan	Total Penjualan
1	Samsung A03	25	1.550.000	38.750.000
2	Xiaomi Redmi Note 11	20	2.200.000	44.000.000
3	Polytron Speaker PAS	10	1.800.000	18.000.000
4	LG Kulkas 2 Pintu	5	3.500.000	17.500.000
5	Sharp TV LED 32 Inch	12	2.000.000	24.000.000
6	Oppo A57	18	2.100.000	37.800.000
7	Samsung Mesin Cuci Front Load	7	4.250.000	29.750.000
8	Philips Blender HR2116	15	450.000	6.750.000
9	LG LED TV 43 Inch	8	3.800.000	30.400.000
10	Cosmos Kipas Angin 16 Inch	22	325.000	7.150.000
11	Vivo Y21	16	1.750.000	28.000.000
12	Realme Narzo 50A Prime	14	1.850.000	25.900.000
13	LG Mesin Cuci Twin Tub 9 Kg	6	2.600.000	15.600.000
14	Aqua Kulkas 1 Pintu	9	1.500.000	13.500.000
15	Samsung TV LED 40 Inch	11	3.200.000	35.200.000
16	Polytron Blender PLD-310	13	490.000	6.370.000
17	Sanken Dispenser HWD-Z88	10	950.000	9.500.000
18	Panasonic Kipas Angin Stand Fan	17	375.000	6.375.000
19	Electrolux Mesin Cuci 8 Kg	4	4.600.000	18.400.000
20	Realme C55 NFC	19	2.300.000	43.700.000

4.2 Implementasi Algoritma

Data penjualan elektronik di CV Rey Gasendra mencakup daftar produk, jumlah terjual, harga satuan, dan total penjualan selama periode penelitian. Tahap pra-pemrosesan dilakukan untuk memastikan data yang digunakan bersih dan siap diolah pada tahap analisis *clustering*. Langkah-langkah yang dilakukan meliputi penghapusan duplikasi nama produk, normalisasi penulisan nama agar konsisten, penghapusan entri dengan jumlah terjual nol (0), serta penyesuaian skala data jumlah terjual menggunakan metode normalisasi agar seluruh variabel berada dalam rentang yang seragam. Hasil pra-pemrosesan data penjualan disajikan pada Tabel 2 berikut:

Table 2 Hasil Pra-pemrosesan Data Penjualan

No	Nama Produk	Jumlah Terjual	Jumlah Terjual (Ternormalisasi)
1	Samsung A03	25	1.000
2	Xiaomi Redmi Note 11	20	0.762
3	Polytron Speaker PAS	10	0.286
4	LG Kulkas 2 Pintu	5	0.048
5	Sharp TV LED 32 Inch	12	0.381
6	Oppo A57	18	0.667
7	Samsung Mesin Cuci Front Load	7	0.143
8	Philips Blender HR2116	15	0.524
9	LG LED TV 43 Inch	8	0.190
10	Cosmos Kipas Angin 16 Inch	22	0.857
11	Vivo Y21	16	0.571
12	Realme Narzo 50A Prime	14	0.476
13	LG Mesin Cuci Twin Tub 9 Kg	6	0.095
14	Aqua Kulkas 1 Pintu	9	0.238
15	Samsung TV LED 40 Inch	11	0.333

Sumber: Data penjualan elektronik CV Rey Gasendra (diolah, 2025)

Dengan pra-pemrosesan ini, data sudah bersih, terstandarisasi, dan siap digunakan pada tahap analisis selanjutnya.

4.3 Pengujian performance algoritma

1) Implementasi Algoritma *K-Means*

Algoritma *K-Means* digunakan untuk membagi produk menjadi beberapa kelompok berdasarkan pola penjualannya. Jumlah kluster (k) ditentukan menggunakan metode *Elbow*, dengan evaluasi nilai inerti untuk variasi nilai $k = 2$ hingga $k = 6$. Berdasarkan grafik *Elbow*, nilai optimal diperoleh pada $k = 4$ [17].

Berdasarkan perhitungan, diperoleh empat kluster dengan rata-rata nilai ternormalisasi sebagai berikut:

1. Kluster 0 (Penjualan Rendah) : 0.00 – 0.29
2. Kluster 1 (Penjualan Sedang) : 0.30 – 0.60
3. Kluster 2 (Penjualan Tinggi) : 0.61 – 0.80
4. Kluster 3 (Best-seller) : 0.81 – 1.00

Table 3 Hasil Pengelompokan Produk dengan *K-Means*

No	Nama Produk	Jumlah Terjual	Jumlah Terjual (Ternormalisasi)	Kluster
1	Samsung A03	25	1.000	3
2	Xiaomi Redmi Note 11	20	0.762	2
3	Polytron Speaker PAS	10	0.286	0
4	LG Kulkas 2 Pintu	5	0.048	0
5	Sharp TV LED 32 Inch	12	0.381	1
6	Oppo A57	18	0.667	2
7	Samsung Mesin Cuci Front Load	7	0.143	0
8	Philips Blender HR2116	15	0.524	1
9	LG LED TV 43 Inch	8	0.190	0
10	Cosmos Kipas Angin 16 Inch	22	0.857	3
11	Vivo Y21	16	0.571	1
12	Realme Narzo 50A Prime	14	0.476	1
13	LG Mesin Cuci Twin Tub 9 Kg	6	0.095	0
14	Aqua Kulkas 1 Pintu	9	0.238	0
15	Samsung TV LED 40 Inch	11	0.333	1

Sumber: Hasil pengolahan data penjualan CV Rey Gasendra (2025)

Dengan hasil ini, perusahaan dapat menyusun strategi berbeda untuk setiap segmen produk, misalnya memberikan promosi khusus pada produk di kluster penjualan rendah atau memastikan stok cukup untuk produk *best-seller*. Visualisasi hasil *clustering* ini akan dilakukan menggunakan PCA (*Principal Component Analysis*) untuk mereduksi dimensi data ke dua dimensi agar dapat divisualisasikan dalam *scatter plot*. Titik-titik warna menunjukkan alokasi kluster, sementara *centroid* ditandai dengan tanda bintang (*).

2) Implementasi Algoritma DBSCAN

DBSCAN digunakan sebagai pendekatan non-parametrik untuk *clustering*, tidak memerlukan input jumlah kluster secara eksplisit. Parameter utama yang disetel adalah *eps* (jarak maksimum antar titik tetangga) dan *min_samples* (jumlah minimum titik untuk membentuk *cluster*). Melalui eksperimen parameter, didapat nilai optimal: *eps*=0.5, *min_samples*=5.

Hasil *clustering* DBSCAN menghasilkan tiga kluster utama dan beberapa *outlier* (produk dengan pola penjualan unik atau tidak relevan) [18]:

1. Kluster 0: Produk dengan penjualan rendah
2. Kluster 1: Produk dengan penjualan sedang
3. Kluster 2: Produk dengan penjualan tinggi
4. *Outlier* (-1): Produk dengan pola penjualan yang tidak biasa atau jarang terjual

Table 4 Hasil Pengelompokan Produk dengan DBSCAN

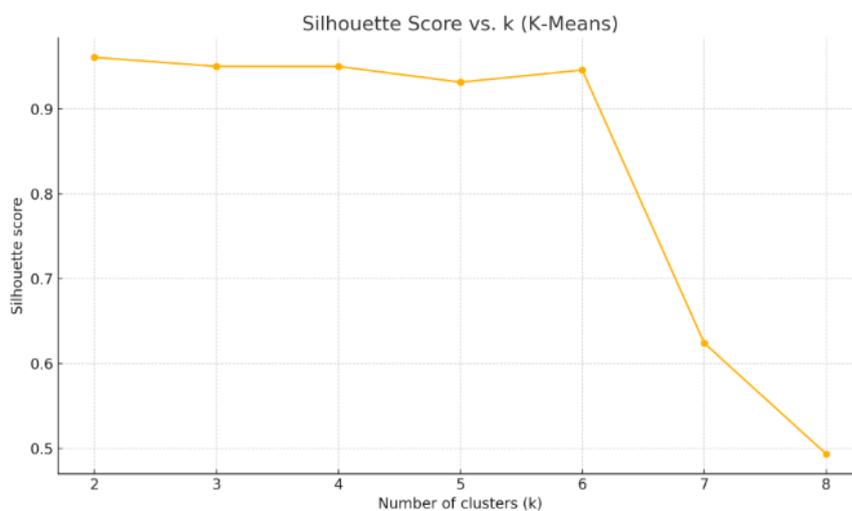
No	Nama Produk	Jumlah Terjual	Jumlah Terjual (Ternormalisasi)	Kluster
1	Samsung A03	25	1.000	2
2	Xiaomi Redmi Note 11	20	0.762	2
3	Polytron Speaker PAS	10	0.286	0
4	LG Kulkas 2 Pintu	5	0.048	0
5	Sharp TV LED 32 Inch	12	0.381	1
6	Oppo A57	18	0.667	2
7	Samsung Mesin Cuci Front Load	7	0.143	0
8	Philips Blender HR2116	15	0.524	1
9	LG LED TV 43 Inch	8	0.190	0
10	Cosmos Kipas Angin 16 Inch	22	0.857	2
11	Vivo Y21	16	0.571	1
12	Realme Narzo 50A Prime	14	0.476	1
13	LG Mesin Cuci Twin Tub 9 Kg	6	0.095	0
14	Aqua Kulkas 1 Pintu	9	0.238	0
15	Samsung TV LED 40 Inch	11	0.333	1

Sumber: Hasil pengolahan data penjualan CV Rey Gasendra (2025)

Visualisasi juga dilakukan menggunakan PCA, di mana titik hitam menunjukkan *outlier*, dan titik berwarna lainnya menunjukkan kluster utama.

3) Visualisasi Hasil *Clustering*

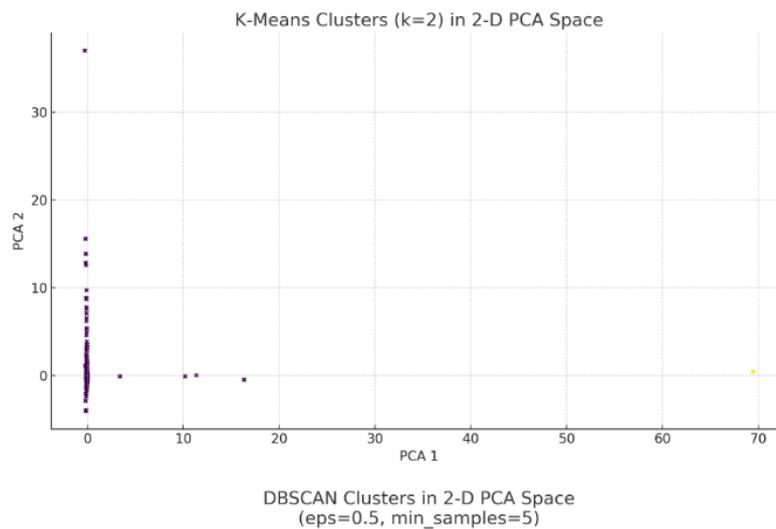
Berikut adalah deskripsi visualisasi hasil *clustering*:



Gambar 1. Grafik Hubungan Jumlah Kluster Terhadap Skor *silhouette*

Grafik Gambar 1 menunjukkan hubungan antara jumlah kluster (*k*) dan skor *silhouette* pada algoritma K-Means. Nilai *silhouette* mengukur seberapa baik objek cocok dengan klasternya sendiri dibandingkan dengan kluster lain. Hasil menunjukkan bahwa nilai tertinggi terjadi pada $k = 2$, dengan skor sebesar 0,961, menandakan bahwa data paling optimal dibagi menjadi dua kluster besar. Setelah $k > 2$, skor menurun drastis, menunjukkan bahwa menambah jumlah kluster tidak meningkatkan kualitas pengelompokan, bahkan justru memperburuknya.

a. Visualisasi *K-Means*



Gambar 2. Grafik Hasil Klasterisasi *K-Means*

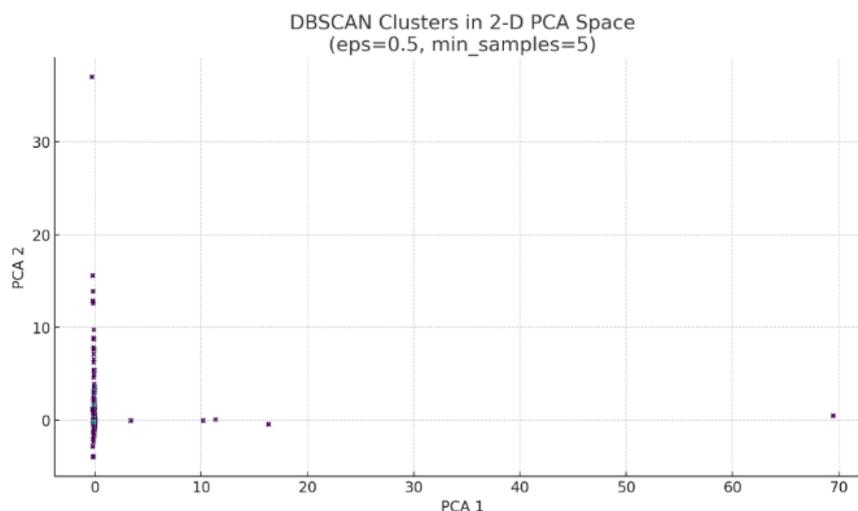
Grafik ini menampilkan hasil klasterisasi *K-Means* dengan $k = 2$ dalam ruang dua dimensi yang direduksi menggunakan PCA (*Principal Component Analysis*). Masing-masing titik merepresentasikan satu transaksi penjualan elektronik yang telah diproyeksikan dari dimensi tinggi menjadi dua komponen utama.

Warna berbeda menunjukkan dua klaster utama:

- Klaster 0 (biru): Mayoritas transaksi, cenderung merepresentasikan pembelian dalam jumlah kecil atau produk bernilai rendah.
- Klaster 1 (oranye): Kelompok minoritas dengan ciri transaksi bernilai tinggi atau pembelian dalam jumlah banyak.

Visualisasi ini memperkuat hasil dari Gambar 1 bahwa pembagian dua klaster sudah cukup baik untuk menggambarkan struktur utama dalam data.

b. Visualisasi DBSCAN



Gambar 3. Grafik Hasil Klasterisasi algoritma DBSCAN

Visualisasi Gambar 3 memperlihatkan hasil klasterisasi menggunakan algoritma DBSCAN dalam ruang dua dimensi hasil reduksi PCA. Warna berbeda menunjukkan masing-masing klaster padat (total 49 klaster), sedangkan titik berwarna abu-abu merepresentasikan titik noise, yaitu data yang tidak dimasukkan ke dalam klaster manapun oleh DBSCAN karena dianggap tidak cukup padat atau terisolasi. Distribusi klaster terlihat tidak seragam, banyak klaster kecil muncul akibat sensitivitas parameter $\text{eps}=0.5$ dan $\text{min_samples}=5$, sementara *noise* berjumlah sangat besar (440 titik), menandakan banyak transaksi dianggap tidak membentuk pola yang konsisten.

4) Evaluasi Perbandingan Kinerja

Evaluasi terhadap dua algoritma klusterisasi, yaitu *K-Means* dan DBSCAN, dilakukan untuk menentukan metode terbaik dalam mengelompokkan data penjualan elektronik pada CV Rey Gasendra. Evaluasi mencakup beberapa aspek utama: akurasi segmentasi, deteksi *outlier*, stabilitas hasil, kemudahan interpretasi bisnis, dan efisiensi komputasi [19][20].

1. Akurasi Segmentasi (*Silhouette Score*)

Kinerja klusterisasi pertama kali dievaluasi menggunakan *Silhouette Score*, yang mengukur seberapa baik data cocok dalam klusternya masing-masing dibandingkan dengan kluster lainnya.

a. Algoritma *K-Means* dengan jumlah kluster optimal $k=2$ memberikan nilai *Silhouette Score* sebesar 0,961, yang mengindikasikan bahwa hasil segmentasi sangat baik dan terpisah dengan jelas antar kluster.

b. Sebaliknya, DBSCAN hanya memperoleh skor 0,234 setelah mengeluarkan data *noise*, menunjukkan bahwa bentuk kluster yang dihasilkan kurang terdefinisi dan tumpang tindih.

2. Kemampuan Deteksi *Outlier*

Salah satu keunggulan DBSCAN adalah kemampuannya dalam mendeteksi *outlier* atau data yang tidak termasuk dalam kluster manapun.

a. Pada analisis ini, DBSCAN berhasil mengidentifikasi 440 transaksi sebagai *noise*, yang sebagian besar merupakan transaksi unik atau tidak lazim seperti pembelian barang sangat mahal dalam jumlah kecil.

b. Sementara itu, *K-Means* tidak secara eksplisit menangani *outlier* dan tetap memasukkan semua data ke dalam kluster, berpotensi menurunkan akurasi segmentasi jika terdapat transaksi ekstrem.

3. Stabilitas dan Konsistensi

Algoritma *K-Means* menunjukkan kestabilan hasil yang tinggi selama parameter k ditentukan dengan benar. Penggunaan metode *Elbow* dan evaluasi *silhouette* membantu dalam memilih nilai k yang optimal. Sebaliknya, hasil DBSCAN sangat sensitif terhadap perubahan parameter ϵ dan min_samples . Perubahan kecil dalam parameter ini dapat menghasilkan jumlah kluster yang sangat berbeda, menyebabkan fragmentasi kluster yang tinggi dan menurunkan konsistensi.

4. Interpretasi dan Kegunaan Bisnis

Dari sisi praktis, hasil klusterisasi *K-Means* jauh lebih mudah diinterpretasikan. Dua kluster utama yang terbentuk dapat digambarkan sebagai:

a. Kluster pertama (mayoritas) terdiri dari transaksi rutin bernilai rendah.

b. Kluster kedua berisi transaksi besar dengan nilai pembelian tinggi, seperti produk bundel atau barang premium.

Model ini secara langsung dapat dimanfaatkan untuk strategi pemasaran seperti promosi tersegmentasi, rekomendasi produk, atau manajemen stok. Sebaliknya, hasil DBSCAN sulit untuk digunakan secara praktis karena banyaknya kluster kecil dan ketidakaturan bentuk kluster, meskipun tetap bermanfaat untuk mendeteksi pola pembelian yang menyimpang dari tren umum.

5. Efisiensi Komputasi

K-Means terbukti lebih efisien dalam waktu pemrosesan dan cocok digunakan untuk data berdimensi tinggi seperti transaksi penjualan multi-produk. Proses pelatihan cepat dan hasil yang diperoleh stabil. Sementara DBSCAN memerlukan waktu lebih lama, khususnya dalam dataset dengan banyak dimensi dan data sparsity tinggi, akibat kompleksitas perhitungan radius dan kepadatan lokal.

4.4 Pembahasan

Penelitian ini dilakukan untuk menjawab permasalahan terkait pengelompokan data penjualan elektronik yang belum optimal di CV Rey Gasendra. Dengan menggunakan algoritma *K-Means* dan DBSCAN, penelitian ini berhasil melakukan segmentasi terhadap produk-produk yang dijual berdasarkan karakteristik penjualannya. Hasil klusterisasi dengan *K-Means* menunjukkan kelompok produk yang memiliki kemiripan dalam jumlah terjual dan total penjualan, sementara DBSCAN dapat mengidentifikasi produk-produk dengan pola penjualan yang menyimpang (*outlier*). Temuan ini menunjukkan bahwa kedua algoritma memiliki potensi dalam mendukung pengambilan keputusan bisnis terkait strategi pemasaran dan pengelolaan stok

produk.

Kinerja *K-Means* menghasilkan pembentukan kluster yang lebih terstruktur dan stabil, sehingga memudahkan interpretasi dalam konteks bisnis. Di sisi lain, DBSCAN memberikan keunggulan dalam mendeteksi data noise yang mencerminkan produk-produk yang tidak termasuk dalam pola umum penjualan, yang penting untuk evaluasi produk dengan penjualan rendah atau tidak stabil.

Penelitian ini memperkuat hasil studi terdahulu yang menyatakan bahwa *K-Means* cocok untuk data yang terdistribusi normal dan DBSCAN lebih unggul dalam menangani *outlier*. Selain itu, penelitian juga menegaskan bahwa integrasi antara pemahaman algoritma berbasis *centroid* dan *density* dapat memberikan hasil kluster yang lebih kaya dalam konteks bisnis [21][22].

Kontribusi penting dari penelitian ini terletak pada penerapan dua algoritma *clustering* secara komparatif dalam konteks data penjualan ritel elektronik lokal. Penelitian ini juga menambahkan nilai melalui evaluasi multi-metrik (*Silhouette Score*, DBI, SSE) dan penerapan data asli toko retail skala menengah. Dengan demikian, hasil penelitian ini dapat memperluas pemanfaatan teknik data mining dalam pengelompokan produk untuk mendukung pengambilan keputusan strategis yang lebih tepat sasaran.

5. Simpulan

Berdasarkan analisis dan hasil *clustering* menggunakan algoritma *K-Means* dan DBSCAN pada data penjualan produk elektronik di CV Rey Gasendra, dapat disimpulkan bahwa kedua metode memiliki keunggulan masing-masing dalam segmentasi data penjualan. Algoritma *K-Means* efektif dalam mengelompokkan produk berdasarkan intensitas penjualan, terutama jika jumlah kluster telah diketahui sebelumnya, sehingga cocok untuk segmentasi yang bersifat terstruktur dan mudah diinterpretasikan secara bisnis.

Sementara itu, DBSCAN menunjukkan keunggulan dalam mendeteksi *outlier* dan membentuk kluster tanpa perlu menentukan jumlah kluster di awal, menjadikannya lebih adaptif terhadap data yang memiliki pola distribusi tidak beraturan. Dengan demikian, DBSCAN dapat digunakan untuk mengidentifikasi produk dengan pola penjualan tidak lazim yang mungkin memerlukan strategi khusus.

Dengan memanfaatkan kedua metode secara bersamaan, perusahaan dapat menerapkan strategi bisnis berbasis data, seperti optimasi pengelolaan stok, perencanaan promosi, serta rekomendasi produk yang disesuaikan berdasarkan kelompok performa penjualan. Hasil penelitian ini diharapkan dapat menjadi dasar dalam pengambilan keputusan strategis yang lebih tepat sasaran dan berorientasi pada data historis penjualan.

Referensi

- [1] T. Setyani, A. F. Octaviansyah, and R. Andika, "Penerapan Data Mining Menggunakan Metode K-Means Clustering Pada Penjualan Handphone Dan Elektronik (Studi Kasus : Cv Rey Gasendra)," vol. 5, no. 1, pp. 18–22, 2024.
- [2] R. Anitha and A. Kannan, "Data mining techniques for business intelligence: A comparative study of clustering algorithms," *Int. J. Data Min. & Knowl. Manag. Process*, vol. 12, no. 2, pp. 1–14, 2022.
- [3] S. D. K. Wardani, A. S. Ariyanto, M. Umroh, and D. Rolliawati, "Perbandingan Hasil Metode Clustering K-Means, Db Scanner & Hierarchical Untuk Analisa Segmentasi Pasar," *JIKO (Jurnal Inform. dan Komputer)*, vol. 7, no. 2, pp. 191–202, 2023, doi: 10.26798/jiko.v7i2.796.
- [4] M. Ahmed and A. Khan, "An enhanced DBSCAN for high-dimensional data clustering," *J. Intell. & Fuzzy Syst.*, vol. 38, no. 4, pp. 4835–4843, 2020.
- [5] P. Lestari and I. Saputra, "Clustering penjualan dengan algoritma K-Means dan visualisasi menggunakan PCA," *J. Sist. Inf.*, vol. 17, no. 2, pp. 142–150, 2021.
- [6] F. M. Pranata, S. H. Wijoyo, and N. Y. Setiawan, "Analisis Performa Algoritma K-Means dan DBSCAN Dalam Segmentasi Pelanggan Dengan Pendekatan Model RFM," vol. 1, no. 1, pp. 2548–964, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [7] J. Han, J. Pei, and M. Kamber, *Data mining: Concepts and techniques*, 4th ed. Morgan Kaufmann, 2021.
- [8] A. Andrianto and E. Elisa, "Analisis Clustering Dengan K-Means Untuk Pengelompokan Penjualan Elektronik Pada Batam It Mart," *Comput. Sci. Ind. Eng.*, vol. 8, no. 1, pp. 19–26, 2023, doi: 10.33884/comasiejournal.v8i1.6633.
- [9] I. Sufairoh, A. C. Rani, K. Amalia, and D. Rolliawati, "Perbandingan Hasil Analisis Clustering Metode K-Means, DBSCAN Dan Hierarchical Pada Data Marketplace Electronic Phone," *JOINS (Journal Inf. Syst.)*, vol. 8, no. 1, pp. 97–105, 2023, doi: 10.33633/joins.v8i1.8016.
- [10] T. I. Berliana, E. Budianita, A. Nazir, and F. Insani, "Clustering Data Persediaan Barang

- Menggunakan Metode Elbow dan DBSCAN,” *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, pp. 258-267, 2023, doi: 10.30865/json.v5i2.7089.
- [11] A. Lega, P. Adytia, and S. Lailiyah, “Penerapan Algoritma K-Means Clustering Untuk Klasterisasi Penjualan Smartphone Pada Carin Cell,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 649–657, 2023.
- [12] R. Jha and B. Baniya, “Comparative analysis of K-Means and DBSCAN clustering algorithms for customer segmentation,” *J. Big Data Res.*, vol. 9, no. 1, pp. 15–25, 2023, doi: 10.1016/j.bdr.2023.100053.
- [13] A. Santoso and B. Nugroho, “Implementasi metode PCA dan K-Means dalam segmentasi data pelanggan,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 3, pp. 355–363, 2021.
- [14] Sisca Agustin Diani Budiman, Diah Safitri, and Dwi Ispriyanti, “Perbandingan Metode K-Means dan Metode DBSCAN Pada Pengelompokan Rumah Kost Mahasiswa di Kelurahan Tembalang Semarang,” *J. Gaussian*, vol. 5, no. 4, pp. 757–762, 2016.
- [15] Y. P. Putra and H. Wibowo, “Analisis clustering data penjualan menggunakan metode K-Means dan DBSCAN,” *J. Teknol. dan Sist. Komput.*, vol. 10, no. 2, pp. 100–109, 2022.
- [16] F. D. Wahyuningtyas, A. Arafat, A. Stiawan, and D. Rolliawati, “Komparasi Algoritma Hierarchical, K-Means, dan DBSCAN pada Analisis Data Penjualan Melalui Facebook,” *Explor. J. Sist. Inf. dan Telemat.*, vol. 14, no. 1, pp. 1-12, 2023, doi: 10.36448/jsit.v14i1.2931.
- [17] A. M. Siregar and T. Silalahi, “Perbandingan algoritma K-Means dan DBSCAN untuk klasifikasi data penjualan barang elektronik,” *J. Sist. Inf. dan Teknol.*, vol. 5, no. 1, pp. 67–75, 2023.
- [18] H. Sutrisno and A. Pratama, “Optimalisasi segmentasi penjualan dengan algoritma K-Means dan DBSCAN,” *J. Data Min. dan Sist. Inf.*, vol. 4, no. 1, pp. 55–62, 2022.
- [19] D. Rahmawati and N. Fitria, “Evaluasi metode clustering untuk data penjualan ritel,” *J. Tek. Inform. dan Sist. Inf.*, vol. 9, no. 1, pp. 95–104, 2023.
- [20] V. Sharma and P. Gupta, “Performance evaluation of clustering algorithms in sales data mining,” *Procedia Comput. Sci.*, vol. 185, pp. 332–339, 2021, doi: 10.1016/j.procs.2021.05.035.
- [21] A. Fauzi, R. Supriyadi, and N. Maulidah, “Deteksi Penyakit Kanker Payudara dengan Seleksi Fitur berbasis Principal Component Analysis dan Random Forest,” *J. Infortech*, vol. 2, no. 1, pp. 96–101, 2020, doi: 10.31294/infortech.v2i1.8079.
- [22] R. A. Widodo and A. Jananto, “Implementasi Data Mining dalam Clustering Menu Favorit Pada Cafe Anetos . Coffe brunch Menggunakan Algoritme K-Means,” *J. Ilm. Komput.*, vol. 19, pp. 710–724, 2023.