



Jurnal Komputasi

## Perbandingan Algoritma Clustering K-Means dan DBSCAN untuk Segmentasi Outlet Ritel Berdasarkan Pola Transaksi Penjualan

<sup>1,\*</sup>Nurul Azwanti, <sup>2</sup>Yulia

<sup>1</sup> Sistem Informasi, Universitas Nagoya Indonesia, Jl. Prambanan No.99 Jodoh, Batam, Indonesia

<sup>2</sup> Teknik Informatika, Universitas Islam Sumatera Barat, Jl. By Pass Km.22, Padang, Indonesia

**Abstrak** — Perkembangan sektor ritel yang pesat menghasilkan volume data transaksi yang besar dan berpotensi dimanfaatkan sebagai dasar pengambilan keputusan bisnis yang lebih efektif. Namun, pemanfaatan data tersebut masih belum optimal, khususnya dalam melakukan segmentasi outlet berdasarkan pola transaksi penjualan. Penelitian ini bertujuan untuk melakukan segmentasi outlet ritel di Kabupaten Kampar menggunakan algoritma K-Means dan DBSCAN, serta membandingkan kinerja kedua metode dalam mengelompokkan data transaksi. Data yang digunakan merupakan data transaksi penjualan yang telah melalui tahap pembersihan data, agregasi dari level transaksi ke level outlet, serta normalisasi untuk menyamakan skala antar variabel. Setiap outlet direpresentasikan berdasarkan total transaksi, total quantity, total penjualan, dan rata-rata transaksi, sehingga dapat dianalisis secara komprehensif. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu menghasilkan tiga cluster utama, yaitu outlet kecil, menengah, dan besar, dengan kualitas clustering yang sangat baik yang ditunjukkan oleh nilai Silhouette Score sebesar 0,960. Sementara itu, algoritma DBSCAN hanya menghasilkan satu cluster utama dan sejumlah data yang teridentifikasi sebagai noise atau outlier, sehingga kurang optimal dalam membentuk segmentasi pada dataset yang cenderung homogen. Meskipun demikian, DBSCAN memiliki keunggulan dalam mendeteksi outlet dengan karakteristik ekstrem yang memiliki kontribusi signifikan terhadap penjualan. Berdasarkan hasil perbandingan, K-Means lebih efektif digunakan sebagai metode utama dalam segmentasi outlet, sedangkan DBSCAN berperan sebagai metode pendukung dalam identifikasi outlier. Hasil penelitian ini diharapkan dapat mendukung penyusunan strategi pemasaran, distribusi, dan pengelolaan outlet yang lebih tepat sasaran berbasis analisis data.

**Kata Kunci:** Clustering; K-Means; DBSCAN; Segmentasi Outlet; Data Transaksi Ritel.

**Abstract** — The rapid growth of the retail sector has led to a significant increase in transaction data, which has great potential to support more effective business decision-making. However, the utilization of this data remains suboptimal, particularly in segmenting retail outlets based on sales transaction patterns. This study aims to segment retail outlets in Kampar Regency using K-Means and DBSCAN algorithms, as well as to compare the performance of both methods in clustering transaction data. The dataset used consists of sales transaction data that has undergone data cleaning, aggregation from transaction level to outlet level, and normalization to ensure consistent variable scales. Each outlet is represented by total transactions, total quantity, total sales, and average transaction value, enabling a comprehensive analysis. The results show that the K-Means algorithm successfully forms three main clusters, namely small, medium, and large outlets, with excellent clustering quality indicated by a Silhouette Score of 0.960. In contrast, the DBSCAN algorithm produces only one main cluster and identifies several data points as noise or outliers, making it less effective in forming meaningful segmentation on relatively homogeneous data. Nevertheless, DBSCAN demonstrates strength in detecting outlets with extreme characteristics that significantly contribute to overall sales. Therefore, K-Means is considered more suitable as the primary method for outlet segmentation, while DBSCAN serves as a complementary approach for outlier detection. The findings of this study are expected to support the development of more targeted marketing, distribution, and outlet management strategies based on data-driven insights.

**Keywords:** Clustering; K-Means; DBSCAN; Outlet Segmentation; Retail Transaction Data.

\* Corresponding author :

Nurul Azwanti

Universitas Nagoya Indonesia, Batam, Indonesia

nurulazw21@gmail.com

## 1. PENDAHULUAN

Perkembangan sektor ritel di Indonesia dalam beberapa tahun terakhir menunjukkan tren pertumbuhan yang konsisten dan positif. Hal ini tercermin dari meningkatnya konsumsi rumah tangga serta aktivitas perdagangan yang semakin dinamis. Data Bank Indonesia mencatat bahwa Indeks Penjualan Riil mengalami pertumbuhan sebesar 5,1% pada akhir tahun 2024, sedangkan Badan Pusat Statistik melaporkan konsumsi rumah tangga tumbuh sekitar 5% dan menjadi kontributor utama dalam pertumbuhan ekonomi nasional [1][2]. Fakta ini menegaskan bahwa sektor ritel memiliki peran strategis dalam menopang perekonomian Indonesia. Selain itu, aktivitas ritel yang semakin intensif menghasilkan volume data transaksi yang besar dan terus bertambah setiap harinya. Data transaksi tersebut, apabila diolah dengan tepat, memiliki potensi besar untuk menjadi sumber informasi strategis yang dapat mendukung pengambilan keputusan bisnis yang lebih efektif.

Besarnya volume data transaksi ritel membuka peluang penerapan analisis data modern untuk mengubah data mentah menjadi pengetahuan yang bernilai. Salah satu pendekatan yang relevan adalah *data mining*, khususnya teknik *clustering*, yang bertujuan mengelompokkan data berdasarkan kemiripan karakteristik. *Clustering* telah banyak digunakan dalam berbagai bidang seperti kesehatan, pariwisata, dan transportasi untuk menemukan pola tersembunyi dalam data yang tidak terstruktur [3][4]. Selain itu, penerapan data mining juga telah dimanfaatkan dalam analisis perilaku konsumen dan pengambilan keputusan bisnis, seperti pada analisis pola transaksi dan segmentasi data berbasis association rule maupun klasifikasi [5][6]. Dalam konteks ritel, penerapan *clustering* dapat digunakan untuk melakukan segmentasi outlet berdasarkan pola transaksi penjualan. Segmentasi ini penting karena memungkinkan pelaku usaha untuk merancang strategi pemasaran, distribusi, dan pengelolaan stok secara lebih tepat sasaran, sehingga efisiensi operasional dapat ditingkatkan dan daya saing usaha semakin kuat [7].

Salah satu algoritma *clustering* yang paling populer adalah K-Means. Algoritma ini banyak digunakan karena kesederhanaan konsep, efisiensi komputasi, serta kemampuannya dalam menangani data berukuran besar [8][9]. Namun, K-Means memiliki keterbatasan yang cukup signifikan. Algoritma ini mengasumsikan bahwa cluster berbentuk sferis dan memiliki ukuran yang relatif seragam. Selain itu, K-Means sangat sensitif terhadap *outlier*, sehingga keberadaan data anomali dapat memengaruhi kualitas hasil segmentasi [10]. Pada data transaksi ritel yang cenderung kompleks, heterogen, dan mengandung ketidakteraturan, keterbatasan ini dapat menyebabkan hasil clustering kurang optimal. Sebagai alternatif, DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) menawarkan pendekatan berbasis densitas yang lebih fleksibel. DBSCAN mampu mengidentifikasi cluster dengan bentuk tidak beraturan serta mendeteksi *noise* atau data anomali [11][12]. Keunggulan ini menjadikan DBSCAN lebih sesuai untuk menangani data dunia nyata yang umumnya tidak homogen. Namun demikian, DBSCAN juga memiliki kelemahan, terutama dalam penentuan parameter yang sensitif serta kompleksitas komputasi yang lebih tinggi dibandingkan K-Means [13].

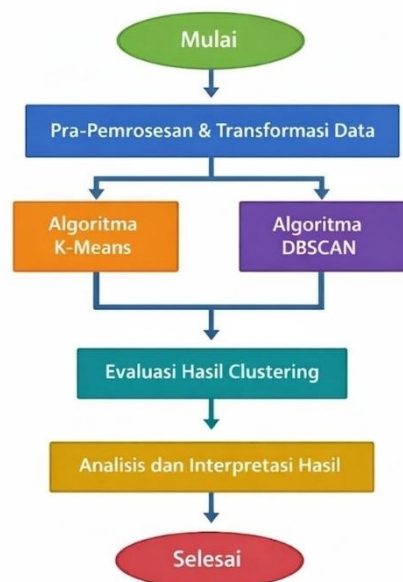
Berbagai penelitian sebelumnya menunjukkan bahwa tidak ada algoritma *clustering* yang selalu unggul dalam semua kondisi. K-Means cenderung memberikan hasil yang baik pada data dengan distribusi sederhana dan homogen, sedangkan DBSCAN lebih unggul pada data yang memiliki *noise* dan pola distribusi kompleks [14][15]. Oleh karena itu, pemilihan algoritma *clustering* harus disesuaikan dengan karakteristik data yang digunakan. Analisis komparatif menjadi penting untuk menentukan metode yang paling sesuai dengan kondisi data tertentu [16]. Dalam konteks Kabupaten Kampar, pemanfaatan data transaksi ritel masih belum optimal. Padahal, dengan segmentasi outlet yang tepat, pelaku usaha dapat merancang strategi pemasaran dan distribusi yang lebih efektif. Tanpa adanya segmentasi yang akurat, strategi yang diterapkan berpotensi tidak efisien dan kurang tepat sasaran [7][17].

Kebaruan penelitian ini terletak pada penerapan serta perbandingan algoritma K-Means dan DBSCAN dalam konteks segmentasi outlet ritel berbasis pola transaksi penjualan di Kabupaten Kampar. Berbeda dengan penelitian sebelumnya yang lebih banyak berfokus pada sektor kesehatan, transportasi, atau pariwisata, penelitian ini menyoroti sektor ritel lokal dengan karakteristik data yang spesifik. Selain membandingkan hasil *clustering*, penelitian ini juga mengevaluasi kemampuan masing-masing algoritma dalam menangani *noise* dan distribusi data yang tidak merata—karakteristik umum dalam data transaksi ritel [15]. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi baik

secara akademis maupun praktis. Secara akademis, penelitian ini memperkaya literatur terkait perbandingan algoritma *clustering* dalam analisis data ritel. Secara praktis, hasil penelitian ini diharapkan dapat membantu pelaku usaha dalam merancang strategi pemasaran, distribusi, dan pengelolaan outlet yang lebih efektif berbasis segmentasi data yang akurat dan berbasis bukti.

## 2. METODOLOGI PENELITIAN

Metodologi penelitian ini dirancang secara sistematis untuk menghasilkan segmentasi outlet ritel yang akurat berdasarkan pola transaksi penjualan. Proses penelitian dilakukan melalui beberapa tahapan utama yang saling terintegrasi, dimulai dari pengumpulan data transaksi, dilanjutkan dengan tahap pra-pemrosesan data yang mencakup pembersihan data, agregasi, dan normalisasi. Data yang telah diproses kemudian dianalisis menggunakan algoritma K-Means dan DBSCAN untuk membentuk kelompok outlet berdasarkan kemiripan karakteristik. Selanjutnya, hasil clustering dievaluasi menggunakan metrik Silhouette Score dan Davies-Bouldin Index guna mengukur kualitas pengelompokan yang dihasilkan. Pendekatan yang digunakan dalam penelitian ini adalah kuantitatif dengan metode eksperimen komparatif, sehingga memungkinkan perbandingan kinerja kedua algoritma dilakukan secara objektif pada dataset yang sama. Alur keseluruhan tahapan penelitian ini disajikan pada Gambar 1.



Gambar 1. Metodologi Penelitian

### 2.1 Pra-pemrosesan dan Transformasi Data

Tahap pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum analisis clustering. Proses ini meliputi pembersihan data dengan menghapus duplikasi, menangani nilai yang hilang, serta menyeragamkan format numerik agar konsisten. Selanjutnya, dilakukan proses agregasi data dari level transaksi menjadi level outlet, mengingat data awal masih berbentuk per faktur. Pada tahap ini dibentuk sejumlah variabel baru yang merepresentasikan karakteristik perilaku outlet, yaitu total transaksi (jumlah faktur), total quantity pembelian, total nilai penjualan, serta rata-rata nilai transaksi per outlet. Variabel-variabel tersebut dipilih karena mampu mencerminkan intensitas dan nilai ekonomi aktivitas pembelian outlet. Setelah proses agregasi, dilakukan normalisasi data menggunakan metode Min-Max Scaling atau standarisasi untuk menghindari dominasi atribut tertentu akibat perbedaan skala nilai, sehingga proses clustering dapat menghasilkan pengelompokan yang lebih optimal.

## 2.2 Penerapan Algoritma K-Means

Algoritma K-Means digunakan untuk mengelompokkan outlet berdasarkan kemiripan pola transaksi yang telah direpresentasikan dalam bentuk fitur numerik. Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow dan Silhouette Score untuk memperoleh jumlah cluster yang paling representatif. Proses clustering dimulai dengan inialisasi centroid secara acak, kemudian dilanjutkan dengan perhitungan jarak antar data menggunakan Euclidean Distance. Setiap data akan dialokasikan ke cluster dengan jarak terdekat, dan centroid diperbarui secara iteratif hingga mencapai kondisi konvergen. Hasil akhir dari algoritma ini berupa kelompok outlet dengan karakteristik transaksi yang homogen dalam satu cluster dan heterogen antar cluster.

## 2.3 Penerapan Algoritma DBSCAN

Algoritma DBSCAN diterapkan sebagai metode clustering berbasis densitas untuk mengatasi keterbatasan K-Means dalam mendeteksi bentuk cluster yang tidak beraturan dan keberadaan data noise. DBSCAN tidak memerlukan penentuan jumlah cluster di awal, melainkan menggunakan parameter epsilon ( $\epsilon$ ) sebagai radius pencarian dan minimum points (MinPts) sebagai jumlah minimum titik untuk membentuk suatu cluster. Proses clustering dilakukan dengan mengidentifikasi core point, border point, dan noise berdasarkan kepadatan data. Cluster kemudian terbentuk dari kumpulan titik yang saling terhubung dalam radius tertentu. Keunggulan utama DBSCAN dalam penelitian ini adalah kemampuannya mendeteksi outlet dengan pola transaksi yang menyimpang sebagai noise.

## 2.4 Evaluasi Hasil Clustering

Evaluasi dilakukan untuk mengukur kualitas hasil clustering yang dihasilkan oleh kedua algoritma. Metrik yang digunakan meliputi Silhouette Score untuk mengukur tingkat kohesi dan separasi cluster, serta Davies-Bouldin Index untuk menilai tingkat kemiripan antar cluster. Selain itu, jumlah cluster yang terbentuk juga dianalisis untuk melihat struktur pengelompokan data. Khusus pada DBSCAN, dilakukan evaluasi terhadap jumlah data yang teridentifikasi sebagai noise untuk menilai kemampuan algoritma dalam mendeteksi anomali. Hasil evaluasi ini digunakan sebagai dasar dalam membandingkan performa kedua algoritma secara komprehensif.

## 2.5 Analisa dan Interpretasi Hasil

Tahap akhir penelitian adalah menganalisis hasil clustering untuk mengidentifikasi segmentasi outlet, misalnya outlet dengan tingkat penjualan tinggi, sedang, dan rendah. Interpretasi dilakukan dengan melihat karakteristik masing-masing cluster berdasarkan variabel utama. Hasil segmentasi ini diharapkan memberikan wawasan strategis bagi pelaku usaha dalam menentukan kebijakan pemasaran, distribusi produk, serta pengelolaan stok barang. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi akademis, tetapi juga nilai praktis dalam pengambilan keputusan berbasis data.

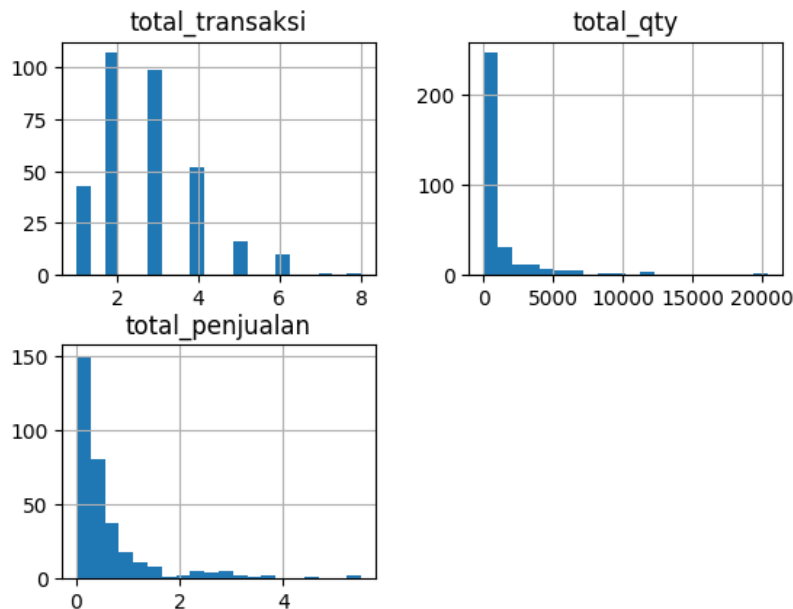
## 3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk melakukan segmentasi outlet ritel di Kabupaten Kampar berdasarkan pola transaksi penjualan menggunakan algoritma K-Means dan DBSCAN. Tahapan analisis dimulai dari pengolahan data, dilanjutkan dengan penerapan algoritma clustering, hingga evaluasi dan interpretasi hasil. Hasil yang diperoleh diharapkan mampu memberikan gambaran karakteristik outlet sebagai dasar pengambilan keputusan bisnis.

### 3.1. Pengolahan dan Penyajian Data

Pada tahap ini dilakukan pengolahan data transaksi penjualan untuk menghasilkan dataset yang siap digunakan dalam analisis clustering. Data awal masih berada pada level transaksi per faktur, sehingga satu outlet dapat memiliki banyak entri. Oleh karena itu, dilakukan transformasi ke level outlet melalui teknik agregasi berdasarkan identitas outlet (CUSTID), sehingga setiap outlet direpresentasikan dalam satu baris data. Proses pengolahan diawali dengan pembersihan data, meliputi penghapusan duplikasi dan penanganan nilai yang tidak lengkap. Berdasarkan hasil observasi, variabel *QTY* memiliki banyak

nilai kosong, sehingga digunakan variabel *FRACT* sebagai representasi jumlah pembelian yang lebih konsisten. Selanjutnya dilakukan konversi tipe data numerik pada variabel yang digunakan untuk memastikan data dapat diproses secara komputasional, kemudian dilakukan agregasi untuk menghasilkan variabel utama yaitu total transaksi, total quantity pembelian, total nilai penjualan, serta rata-rata nilai transaksi per outlet.



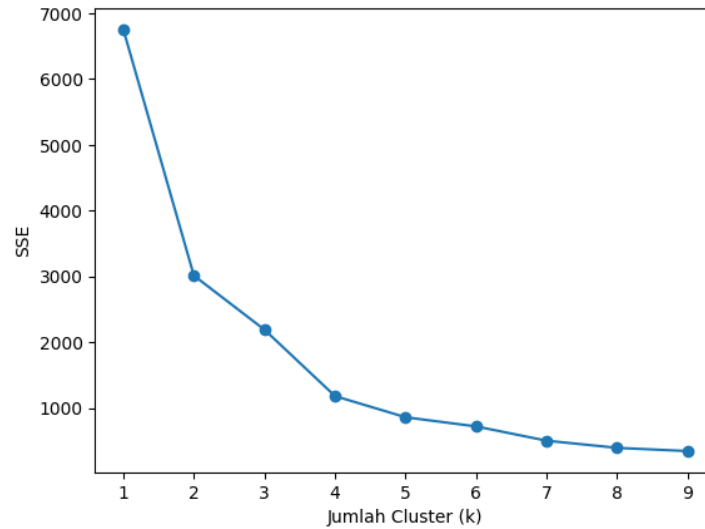
Gambar 2. Distribusi data Transaksi Outlet

Analisis distribusi data dilakukan untuk memahami karakteristik awal dataset. Berdasarkan Gambar 2, distribusi data menunjukkan pola tidak merata (*skewed*), terutama pada variabel *total quantity* dan *total penjualan* yang condong ke kanan. Hal ini mengindikasikan bahwa sebagian besar outlet memiliki nilai transaksi yang relatif rendah, sementara hanya sebagian kecil outlet memiliki nilai transaksi yang tinggi. Kondisi tersebut menunjukkan adanya ketimpangan kontribusi antar outlet dan memperkuat pentingnya segmentasi. Oleh karena itu, dilakukan normalisasi data menggunakan metode *Min-Max Scaling* untuk menyamakan skala antar variabel agar tidak terjadi dominasi dalam proses clustering. Variabel yang dinormalisasi meliputi total transaksi, total quantity, total penjualan, dan rata-rata transaksi, sehingga diperoleh dataset yang telah terstandarisasi dan siap digunakan pada tahap analisis clustering menggunakan algoritma K-Means dan DBSCAN.

### 3.2. Clustering Menggunakan Algoritma K-Means

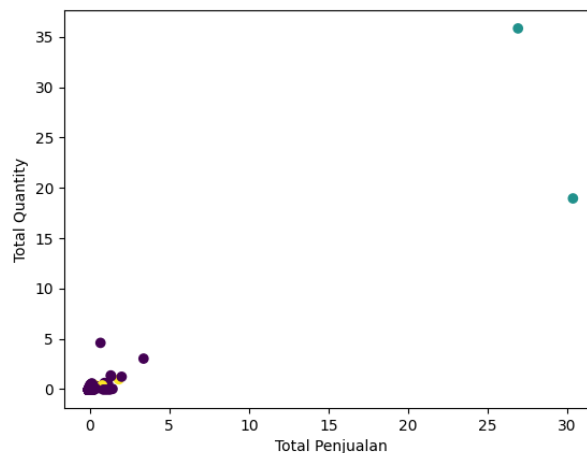
Pada tahap ini dilakukan segmentasi outlet menggunakan algoritma K-Means untuk mengelompokkan outlet berdasarkan karakteristik transaksi yang telah diolah pada tahap sebelumnya. K-Means merupakan metode clustering non-hierarki yang bertujuan membagi data ke dalam beberapa kelompok (*cluster*) berdasarkan kedekatan jarak antar data. Variabel yang digunakan meliputi total transaksi, total quantity, total penjualan, dan rata-rata transaksi per outlet, yang telah dinormalisasi menggunakan *StandardScaler* agar semua variabel berada pada skala yang sama.

Sebelum melakukan clustering, jumlah cluster optimal ditentukan menggunakan metode Elbow dengan menghitung nilai Sum of Squared Errors (SSE) untuk beberapa jumlah cluster.



Gambar 3. Plot Metode Elbow

Berdasarkan hasil plot Elbow pada Gambar 3, terlihat adanya titik siku pada  $k = 3$ , sehingga jumlah cluster yang digunakan adalah 3. Proses clustering kemudian dilakukan dengan K-Means, di mana setiap outlet dialokasikan ke cluster terdekat berdasarkan jarak Euclidean, dan centroid diperbarui secara iteratif hingga konvergen.



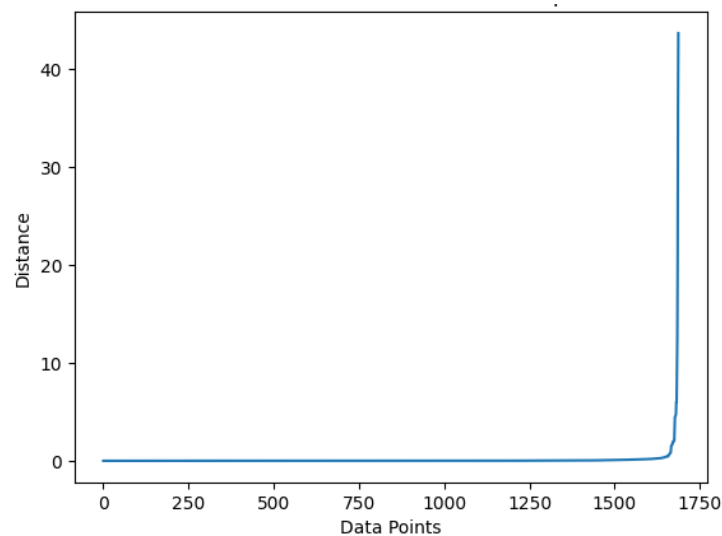
Gambar 4. Hasil Metode Clustering

Hasil clustering divisualisasikan dalam scatter plot dengan sumbu total penjualan dan total quantity, sehingga perbedaan antar cluster terlihat jelas. Analisis karakteristik menunjukkan bahwa Cluster 0 mewakili outlet kecil dengan transaksi dan penjualan rendah, Cluster 1 merupakan outlet sedang dengan aktivitas menengah, sedangkan Cluster 2 terdiri dari outlet besar atau prioritas dengan total transaksi dan penjualan tinggi. Segmentasi ini memberikan dasar yang kuat untuk strategi bisnis, seperti distribusi produk yang lebih intensif pada outlet prioritas, promosi khusus untuk outlet menengah, serta pendekatan efisiensi bagi outlet kecil.

Dengan demikian, hasil clustering K-Means tidak hanya menunjukkan perbedaan karakteristik outlet, tetapi juga memberikan dasar empiris bagi pengambilan keputusan pemasaran, distribusi, dan pengelolaan hubungan dengan outlet secara lebih tepat sasaran.

### 3.3. Clustering Menggunakan Algoritma DBSCAN

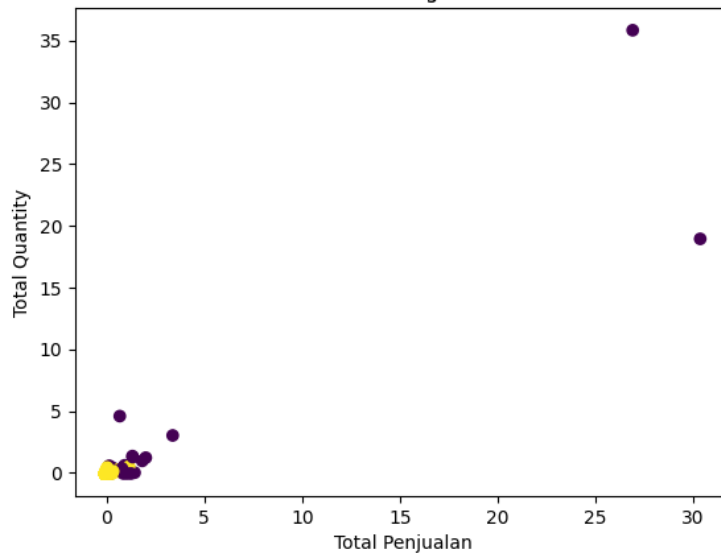
Pada tahap ini dilakukan segmentasi outlet menggunakan algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN) sebagai metode pembandingan terhadap K-Means. DBSCAN merupakan metode clustering berbasis kepadatan yang mengelompokkan data berdasarkan kedekatan antar titik serta mampu mengidentifikasi data yang memiliki karakteristik berbeda secara signifikan sebagai noise (outlier). Algoritma ini menggunakan dua parameter utama, yaitu epsilon ( $\epsilon$ ) sebagai batas jarak maksimum antar titik dan minimum samples (MinPts) sebagai jumlah minimum data untuk membentuk suatu cluster. Dalam DBSCAN, label cluster yang dihasilkan memiliki makna khusus, di mana nilai -1 menunjukkan data yang dikategorikan sebagai noise (outlier), sedangkan nilai selain -1 menunjukkan cluster utama yang terbentuk berdasarkan kepadatan data. Pada penelitian ini, parameter yang digunakan adalah nilai epsilon ( $\epsilon$ ) sebesar **0.8** dan MinPts sebesar **5**, yang diperoleh melalui pendekatan k-distance plot.



Gambar 5. Plot K-Distance untuk Penentuan Epsilon

Penentuan nilai epsilon dilakukan menggunakan metode k-distance plot dengan mengamati titik siku (elbow) pada grafik jarak tetangga terdekat. Berdasarkan hasil visualisasi pada Gambar 5, terlihat adanya peningkatan jarak yang signifikan pada bagian akhir grafik yang menunjukkan batas pemisah antar kelompok data. Nilai epsilon yang dipilih berada pada kisaran yang mampu merepresentasikan kepadatan data secara optimal.

Hasil clustering menunjukkan bahwa sebagian besar outlet tergabung dalam satu cluster utama (cluster 0) sebanyak **1667 outlet (98.76%)**, sedangkan sebagian kecil lainnya teridentifikasi sebagai noise (cluster -1) sebanyak **21 outlet (1.24%)**.



Gambar 6. Hasil Clustering DBSCAN

Visualisasi pada Gambar 6 memperlihatkan bahwa mayoritas outlet memiliki karakteristik transaksi yang relatif seragam dan terkonsentrasi pada nilai rendah, sedangkan sejumlah kecil outlet berada jauh dari kelompok utama sehingga dikategorikan sebagai outlier.

Berdasarkan hasil agregasi, cluster utama (cluster 0) mencerminkan mayoritas outlet dengan jumlah transaksi, total quantity, dan nilai penjualan yang relatif rendah, sehingga termasuk kategori outlet kecil. Sebaliknya, cluster noise (cluster -1) memiliki nilai yang jauh lebih tinggi pada seluruh variabel, sehingga dapat diinterpretasikan sebagai outlet prioritas (key account) dengan kontribusi besar terhadap penjualan. Kondisi ini menunjukkan bahwa DBSCAN kurang optimal dalam membentuk segmentasi utama karena sebagian besar data memiliki kepadatan yang relatif seragam, sehingga hanya menghasilkan satu cluster dominan. Namun, DBSCAN terbukti efektif dalam mengidentifikasi data outlier yang memiliki karakteristik berbeda secara signifikan. Oleh karena itu, dalam penelitian ini K-Means tetap menjadi metode utama untuk segmentasi, sedangkan DBSCAN berfungsi sebagai metode pendukung dalam mendeteksi anomali data.

Tabel 1. Contoh Karakteristik Outlet Berdasarkan Hasil Clustering DBSCAN

Cluster	CUSTID	Total Transaksi	Total Qty	Total Penjualan	Avg Transaksi
0	5100000018	2	0	581.400	290.700
-1	5100000075	10	77	248.233.198	24.823.319

Sebagai ilustrasi, perbedaan karakteristik antar cluster dapat dilihat pada contoh data outlet. Outlet yang termasuk dalam cluster utama (cluster 0) memiliki jumlah transaksi yang rendah, yaitu hanya sebanyak 2 transaksi dengan total penjualan sebesar 581.400. Sebaliknya, outlet yang termasuk dalam kategori noise (cluster -1) menunjukkan aktivitas yang jauh lebih tinggi, dengan 10 transaksi dan total penjualan mencapai 248.233.198. Perbedaan ini menunjukkan adanya kesenjangan yang signifikan antara outlet kecil dan outlet dengan kontribusi besar, sehingga memperkuat peran DBSCAN dalam mengidentifikasi outlet dengan karakteristik ekstrem (outlier).

### 3.4. Perbandingan K-Means dan DBSCAN

Pada tahap ini dilakukan perbandingan antara algoritma K-Means dan DBSCAN untuk mengevaluasi kinerja masing-masing metode dalam melakukan segmentasi outlet. Perbandingan dilakukan berdasarkan jumlah cluster yang terbentuk, distribusi data, serta hasil evaluasi menggunakan Silhouette Score untuk mengukur kualitas pemisahan antar cluster.

Berdasarkan hasil pengolahan data, K-Means dengan jumlah cluster optimal sebanyak 3 mampu membentuk segmentasi yang jelas, yaitu outlet kecil, menengah, dan besar. Hal ini didukung oleh nilai Silhouette Score sebesar **0,960**, yang menunjukkan bahwa cluster yang terbentuk memiliki tingkat

pemisahan yang sangat baik dan antar cluster tidak saling tumpang tindih. Dengan demikian, K-Means dinilai sangat efektif dalam mengelompokkan outlet berdasarkan karakteristik transaksi.

Sementara itu, DBSCAN hanya menghasilkan satu cluster utama dan sejumlah data yang dikategorikan sebagai noise. Kondisi ini menyebabkan Silhouette Score tidak dapat dihitung karena metode evaluasi tersebut memerlukan minimal dua cluster. Hasil ini menunjukkan bahwa DBSCAN tidak mampu membentuk segmentasi yang optimal pada dataset ini, yang disebabkan oleh karakteristik data yang cenderung homogen dan memiliki kepadatan yang seragam pada sebagian besar outlet.

Tabel 2. Perbandingan Hasil Clustering K-Means dan DBSCAN

Metode	Jumlah Cluster	Silhouette Score	Kemampuan Segmentasi	Deteksi Outlier
K-Means	3 cluster	0,960	Sangat baik	Tidak ada
DBSCAN	1 cluster + noise	Tidak tersedia	Kurang optimal	Sangat baik

Dengan demikian, K-Means lebih optimal digunakan sebagai metode utama dalam segmentasi outlet karena mampu menghasilkan kelompok yang jelas dan terstruktur. Di sisi lain, DBSCAN memiliki keunggulan dalam mendeteksi outlier, yaitu outlet dengan nilai transaksi ekstrem yang berperan sebagai *key account*. Oleh karena itu, kombinasi kedua metode memberikan analisis yang lebih komprehensif, di mana K-Means digunakan untuk segmentasi utama dan DBSCAN sebagai pendukung dalam identifikasi anomali untuk mendukung pengambilan keputusan bisnis yang lebih tepat sasaran.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian, proses pengolahan data yang mencakup pembersihan, agregasi, dan normalisasi berhasil menghasilkan dataset yang siap digunakan untuk analisis clustering, di mana transformasi dari level transaksi ke level outlet memungkinkan setiap outlet direpresentasikan secara komprehensif melalui variabel total transaksi, total quantity, total penjualan, dan rata-rata transaksi. Penerapan algoritma K-Means dengan jumlah cluster optimal sebanyak tiga menghasilkan segmentasi outlet yang jelas, yaitu outlet kecil, menengah, dan besar, yang didukung oleh nilai Silhouette Score sebesar 0,960 yang menunjukkan kualitas clustering sangat baik dengan pemisahan antar cluster yang optimal. Sementara itu, penerapan algoritma DBSCAN menunjukkan bahwa sebagian besar outlet tergabung dalam satu cluster utama dan sebagian kecil lainnya teridentifikasi sebagai noise atau outlier, sehingga mengindikasikan bahwa DBSCAN kurang optimal untuk segmentasi pada dataset yang cenderung homogen, namun tetap efektif dalam mendeteksi outlet dengan karakteristik ekstrem. Secara keseluruhan, K-Means lebih sesuai digunakan sebagai metode utama dalam segmentasi outlet, sedangkan DBSCAN berperan sebagai metode pendukung untuk mengidentifikasi outlier, sehingga hasil penelitian ini dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan bisnis, khususnya dalam penyusunan strategi pemasaran, distribusi, dan pengelolaan outlet yang lebih efektif dan tepat sasaran.

#### DAFTAR PUSTAKA

- [1] Bank Indonesia, "Survei Penjualan Eceran Desember 2024: Penjualan Eceran Diprakirakan Meningkat." 2025. [Online]. Available: [https://www.bi.go.id/id/publikasi/ruang-media/news-release/Pages/sp\\_270525.aspx](https://www.bi.go.id/id/publikasi/ruang-media/news-release/Pages/sp_270525.aspx)
- [2] Badan Pusat Statistik, "Konsumsi rumah tangga tumbuh 4,97 persen pada triwulan II 2025." 2025. [Online]. Available: <https://www.antaranews.com/berita/5016301/bps-konsumsi-rumah-tangga-tumbuh-497-persen-pada-triwulan-ii-2025>
- [3] S. Aryal, "Clustering Techniques in Data Mining," *arXiv*, vol. 2024, no. 1, 2024, doi: 10.48550/arXiv.2401.11402.

- [4] E. Zafarani, “Applications of Clustering in Various Domains,” *arXiv*, vol. 2021, no. 1, 2021, doi: 10.48550/arXiv.2111.08658.
- [5] N. Azwanti and T. Tukino, “Implementasi Algoritma Apriori sebagai Strategi Pemasaran Produk Harian,” *J. Oper. Syst.*, vol. 1, no. 1, pp. 23–28, 2026.
- [6] A. R. Hakim and N. Azwanti, “Penerapan Logika Fuzzy Sugeno Untuk Menentukan Kelayakan Pengiriman Limbah Barang Berbahaya dan Beracun,” *J. Oper. Syst.*, vol. 1, no. 1, pp. 10–22, 2026.
- [7] I. Sufairoh, “Clustering for Retail Data Analysis,” *JOINS*, vol. 8, no. 1, 2023, doi: 10.33633/joins.v8i1.8016.
- [8] D. Putra, “K-Means Clustering Performance Analysis,” *JAIC*, vol. 8, no. 2, 2024, doi: 10.30871/jaic.v8i2.8671.
- [9] V. Baligodugula, “Efficiency of K-Means Algorithm,” *arXiv*, vol. 2025, no. 1, 2025, doi: 10.48550/arXiv.2503.23215.
- [10] I. Doroshenko, “Limitations of K-Means Clustering,” *SWorldJournal*, vol. 23, no. 1, 2024, doi: 10.30888/2663-5712.2024-23-00-038.
- [11] A. Jain, “DBSCAN Clustering Analysis,” in *RAMSITA*, 2025, vol. 2025, no. 1. doi: 10.2991/978-94-6463-716-8\_18.
- [12] V. Wulandari, “DBSCAN Advantages in Noise Detection,” *IJATIS*, vol. 1, no. 1, 2023, doi: 10.57152/ijatiss.v1i1.1135.
- [13] B. Kremers, “DBSCAN Parameter Sensitivity,” *arXiv*, vol. 2021, no. 1, 2021, doi: 10.48550/arXiv.2111.12559.
- [14] A. Ramadhan, “Comparative Study K-Means vs DBSCAN,” *JUTIF*, vol. 6, no. 3, 2025, doi: 10.52436/1.jutif.2025.6.3.4602.
- [15] M. Shodikin, “Clustering Comparison in Real Data,” *MALCOM*, vol. 5, no. 4, 2025, doi: 10.57152/malcom.v5i4.2218.
- [16] F. Salman, “Clustering Selection Based on Data,” *Brilliance*, vol. 3, no. 2, 2023, doi: 10.47709/brilliance.v3i2.3373.
- [17] M. Zuhdi, “Data-driven Decision in Retail,” *JUTIF*, vol. 6, no. 1, 2025, doi: 10.52436/1.jutif.2025.6.1.4183.