

Product Sales Optimization with DBSCAN Clustering Algorithm on e-commerce Dataset

Ahmad Rafael

Department of Informatics, Dinamika Bangsa University, Jambi, Indonesia

E-mail: ahmadrafael010904@gmail.com

*Corresponding Author

Mochammad Firdaus

Department of Informatics, Dinamika Bangsa University, Jambi, Indonesia

E-mail: firdausmochammad57@gmail.com

Daisy Larisa S

Department of Informatics, Dinamika Bangsa University, Jambi, Indonesia

E-mail: daisylarissa12@gmail.com

Received: 10 January, 2025; Accepted: 16 June, 2026; Published: 31 July, 2026

Abstract: This research aims to optimize product sales strategies in e-commerce businesses by utilizing the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering algorithm. DBSCAN is applied to a customer transaction dataset to group customers into segments based on their purchasing patterns, represented by the Quantity and UnitPrice variables. The clustering results demonstrate that DBSCAN effectively identifies customer segments with distinct characteristics, such as loyalty, price sensitivity, and product preferences. This information is then used to design targeted sales strategies, including personalized product recommendations, relevant promotional offers, and efficient inventory management. Implementing DBSCAN-based optimization strategies is expected to improve sales, profitability, and customer satisfaction. This research highlights the effectiveness of the DBSCAN clustering algorithm as a valuable tool for understanding customer behavior and optimizing sales strategies in e-commerce businesses.

Keywords: E-commerce, Sales Optimization, Clustering, DBSCAN, Customer Segmentation, Marketing Strategies.

I. Introduction

Pertumbuhan pesat e-commerce telah menghadirkan peluang yang belum pernah ada sebelumnya bagi bisnis untuk menjangkau basis pelanggan yang lebih luas dan memperluas pangsa pasar mereka. Namun, hal ini juga membawa tantangan dalam memahami perilaku pelanggan dan menyesuaikan strategi pemasaran untuk secara efektif menargetkan segmen pelanggan yang beragam. Dalam lanskap e-commerce yang kompetitif, bisnis perlu mengadopsi pendekatan berbasis data untuk mengoptimalkan strategi penjualan mereka dan memaksimalkan profitabilitas. [9], [11]

Segmentasi pelanggan memainkan peran penting dalam mencapai tujuan ini. Dengan membagi pelanggan ke dalam kelompok-kelompok dengan karakteristik dan pola pembelian yang serupa, bisnis dapat mempersonalisasi upaya pemasaran mereka dan menawarkan produk dan promosi yang sesuai dengan kebutuhan dan preferensi spesifik setiap segmen. Algoritma clustering menyediakan alat yang ampuh untuk segmentasi pelanggan dengan secara otomatis mengidentifikasi pola dan mengelompokkan pelanggan berdasarkan kesamaan mereka dalam data. [2], [3], [17], [18]

Di antara berbagai algoritma clustering, Density-Based Spatial Clustering of Applications with Noise

(DBSCAN) menonjol karena kemampuannya untuk menemukan cluster dengan bentuk yang berubah-ubah dan menangani data yang berisik secara efektif. Tidak seperti metode clustering tradisional seperti K-means, DBSCAN tidak memerlukan penentuan jumlah cluster sebelumnya dan dapat mengidentifikasi outlier, yang mewakili pelanggan dengan perilaku pembelian yang unik. [1], [5], [15]

Penelitian ini berfokus pada penerapan algoritma clustering DBSCAN pada dataset e-commerce untuk mengelompokkan pelanggan berdasarkan kuantitas pembelian dan harga satuan produk. Segmen pelanggan yang dihasilkan kemudian dianalisis untuk memahami karakteristik mereka dan mengembangkan strategi penjualan yang ditargetkan. Dengan menyesuaikan rekomendasi produk, penawaran promosi, dan manajemen inventaris dengan kebutuhan spesifik setiap segmen, bisnis dapat meningkatkan penjualan, profitabilitas, dan kepuasan pelanggan. [10], [12], [19]

Pendahuluan ini memberikan konteks untuk penelitian, menyoroti pentingnya segmentasi pelanggan dalam e-commerce dan keuntungan menggunakan algoritma clustering DBSCAN untuk tujuan ini. Bagian selanjutnya akan membahas metodologi, hasil, dan diskusi temuan penelitian, beserta implikasinya untuk mengoptimalkan strategi penjualan produk dalam bisnis e-commerce. [1]

2. Research Method

2.1. Experiment Setup

Bagian ini menjelaskan langkah-langkah yang terlibat dalam pengaturan eksperimen untuk penelitian ini, yang meliputi pengumpulan data, persiapan data, penerapan algoritma DBSCAN, dan evaluasi hasil. Dataset yang digunakan dalam penelitian ini adalah data transaksi e-commerce yang mencakup informasi tentang pembelian pelanggan, seperti ID invoice, kode stok produk, deskripsi produk, kuantitas, harga satuan, tanggal transaksi, ID pelanggan, dan negara. Data dikumpulkan dari sumber yang terpercaya dan relevan dengan tujuan penelitian. Selanjutnya, data transaksi dibersihkan dengan menangani nilai yang hilang, data duplikat, dan inkonsistensi data. Fitur yang relevan untuk segmentasi pelanggan, seperti kuantitas pembelian (Quantity) dan harga satuan produk (UnitPrice), dipilih untuk dianalisis lebih lanjut. Data yang dipilih kemudian distandarisasi menggunakan StandardScaler untuk memastikan bahwa setiap fitur memiliki skala yang sama dan mencegah fitur dengan varians yang lebih besar mendominasi proses clustering.

Algoritma DBSCAN diterapkan menggunakan library scikit-learn di Python. Parameter DBSCAN, yaitu eps (radius lingkungan) dan min_samples (jumlah minimum titik dalam lingkungan), diatur secara eksperimental untuk mendapatkan hasil clustering yang optimal. Proses clustering dilakukan dengan memasukkan data yang telah distandarisasi ke dalam algoritma DBSCAN, dan setiap data point kemudian diberi label cluster berdasarkan kepadatannya. Hasil clustering dievaluasi secara visual menggunakan scatter plot untuk mengamati pola cluster dan outlier. Analisis deskriptif dilakukan untuk memahami karakteristik setiap cluster berdasarkan fitur Quantity dan UnitPrice. Informasi dari analisis cluster ini digunakan untuk merumuskan strategi penjualan yang ditargetkan untuk setiap segmen pelanggan.

Selanjutnya, strategi penjualan yang direkomendasikan diimplementasikan dalam lingkungan bisnis e-commerce. Pengujian A/B dapat dilakukan untuk membandingkan efektivitas strategi yang ditargetkan dengan strategi konvensional. Metrik seperti tingkat konversi, nilai transaksi rata-rata, dan kepuasan pelanggan dipantau untuk mengevaluasi dampak strategi optimasi. Pengaturan eksperimen ini dirancang untuk memastikan bahwa penelitian dilakukan secara sistematis dan hasilnya dapat diandalkan. Dengan mengikuti langkah-langkah ini, diharapkan penelitian ini dapat memberikan wawasan yang berharga bagi optimasi penjualan produk di bisnis e-commerce. [2] [1]

2.2. Dataset

Dataset yang digunakan dalam penelitian ini adalah dataset transaksi e-commerce yang diperoleh dari Kaggle, yang berisi data transaksi dunia nyata dari toko ritel online. Dataset ini terdiri dari kumpulan catatan, di mana setiap catatan mewakili satu transaksi, dan mencakup atribut seperti InvoiceNo, yang merupakan pengidentifikasi unik untuk setiap transaksi, StockCode, pengidentifikasi unik untuk setiap produk, dan Description, yang memberikan deskripsi tekstual dari produk. Selain itu, terdapat atribut Quantity yang menunjukkan jumlah unit produk yang dibeli dalam transaksi, InvoiceDate yang mencatat tanggal dan waktu transaksi, UnitPrice yang menunjukkan harga satuan produk, CustomerID sebagai pengidentifikasi unik untuk

setiap pelanggan, dan Country yang mencatat negara tempat tinggal pelanggan. Untuk tujuan penelitian ini, fokus penelitian diarahkan pada dua atribut utama, yaitu Quantity dan UnitPrice, yang dipilih karena keduanya secara langsung mencerminkan perilaku pembelian pelanggan. Atribut Quantity menggambarkan jumlah unit yang dibeli oleh pelanggan dalam satu transaksi, sementara UnitPrice mencerminkan harga satuan produk yang dibeli. Kedua atribut ini dianggap penting untuk mengidentifikasi segmen pelanggan yang berbeda berdasarkan pola pembelian mereka. Fig.1 menunjukkan isi dari dataset.

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HE	6	12/1/2010 8:26	2.55	17850	United Kingdom
536365	71053	WHITE METAL LAM	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEA	8	12/1/2010 8:26	2.75	17850	United Kingdom
536365	84029G	KNITTED UNION FL	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	84029E	RED WOOLLY HOTT	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	22752	SET 7 BABUSHKA N	2	12/1/2010 8:26	7.65	17850	United Kingdom
536365	21730	GLASS STAR FROS	6	12/1/2010 8:26	4.25	17850	United Kingdom
536366	22633	HAND WARMER UNI	6	12/1/2010 8:28	1.85	17850	United Kingdom
536366	22632	HAND WARMER REC	6	12/1/2010 8:28	1.85	17850	United Kingdom
536367	84879	ASSORTED COLOUR	32	12/1/2010 8:34	1.69	13047	United Kingdom
536367	22745	POPPY'S PLAYHOU	6	12/1/2010 8:34	2.1	13047	United Kingdom
536367	22748	POPPY'S PLAYHOU	6	12/1/2010 8:34	2.1	13047	United Kingdom
536367	22749	FELTCRAFT PRINCE	8	12/1/2010 8:34	3.75	13047	United Kingdom
536367	22310	IVORY KNITTED MU	6	12/1/2010 8:34	1.65	13047	United Kingdom
536367	84969	BOX OF 6 ASSORTE	6	12/1/2010 8:34	4.25	13047	United Kingdom
536367	22623	BOX OF VINTAGE J	3	12/1/2010 8:34	4.95	13047	United Kingdom
536367	22622	BOX OF VINTAGE A	2	12/1/2010 8:34	9.95	13047	United Kingdom
536367	21754	HOME BUILDING BL	3	12/1/2010 8:34	5.95	13047	United Kingdom
536367	21755	LOVE BUILDING BL	3	12/1/2010 8:34	5.95	13047	United Kingdom
536367	21777	RECIPE BOX WITH I	4	12/1/2010 8:34	7.95	13047	United Kingdom
536367	48187	DOORMAT NEW EN	4	12/1/2010 8:34	7.95	13047	United Kingdom
536368	22960	JAM MAKING SET Y	6	12/1/2010 8:34	4.25	13047	United Kingdom
536368	22913	RED COAT RACK PJ	3	12/1/2010 8:34	4.95	13047	United Kingdom
536368	22912	YELLOW COAT RAC	3	12/1/2010 8:34	4.95	13047	United Kingdom

Fig.1.Contoh Dataset

2.3. Applying DBSCAN for Customer Segmentation

Bagian ini menjelaskan penerapan algoritma clustering DBSCAN pada dataset e-commerce yang telah diproses sebelumnya untuk segmentasi pelanggan. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) adalah algoritma clustering berbasis kepadatan yang mengelompokkan titik data berdasarkan kepadatannya di ruang fitur. DBSCAN mengidentifikasi cluster sebagai wilayah dengan kepadatan tinggi yang dipisahkan oleh wilayah dengan kepadatan rendah. Berbeda dengan metode berbasis partisi seperti K-means, DBSCAN tidak memerlukan pendefinisian jumlah cluster sebelumnya dan dapat menemukan cluster dengan bentuk yang berubah-ubah.

DBSCAN memiliki dua parameter utama yang memengaruhi proses clustering, yaitu eps (epsilon) yang mendefinisikan radius lingkungan di sekitar titik data, menentukan jarak maksimum antara dua titik agar mereka dianggap sebagai tetangga, dan min_samples yang menentukan jumlah minimum titik data yang diperlukan dalam lingkungan agar suatu titik dianggap sebagai titik inti, yang menjadi pusat pembentukan cluster. Proses clustering dimulai dengan inisialisasi, di mana algoritma memilih secara acak titik data yang belum ditetapkan ke cluster. Algoritma kemudian mencari semua titik data dalam jarak eps dari titik yang dipilih dan memeriksa kepadatan. Jika jumlah tetangga lebih besar dari atau sama dengan min_samples, titik yang dipilih ditetapkan sebagai titik inti, dan cluster baru dibentuk. Selanjutnya, algoritma secara iteratif memperluas cluster dengan memasukkan semua titik yang dapat dijangkau secara langsung dan titik yang dapat dijangkau melalui rantai titik inti. Proses ini berlanjut hingga semua titik yang dapat dijangkau secara kepadatan ditambahkan ke cluster, dan algoritma beralih ke titik data berikutnya yang belum ditetapkan. Titik data yang tidak ditetapkan ke cluster mana pun dianggap sebagai outlier atau titik noise. [4]

Penyetelan parameter eps dan min_samples sangat penting untuk mendapatkan hasil clustering yang bermakna. Berbagai teknik dapat digunakan untuk penyetelan parameter, seperti grid search yang melibatkan eksplorasi rentang nilai untuk kedua parameter, grafik k-distance yang membantu dalam menentukan nilai eps, dan inspeksi visual untuk memvisualisasikan hasil clustering. Dalam penelitian ini, DBSCAN diterapkan pada dataset e-commerce yang telah diproses menggunakan fitur Quantity dan UnitPrice. Nilai parameter eps dan min_samples ditentukan melalui eksperimen dan evaluasi, dengan tujuan untuk mencapai segmentasi pelanggan yang optimal. Cluster yang dihasilkan mewakili segmen pelanggan yang berbeda berdasarkan perilaku pembelian mereka. Segmen-segmen ini kemudian dapat dianalisis untuk memahami karakteristik mereka dan mengembangkan strategi optimasi penjualan yang ditargetkan. Bagian ini memberikan deskripsi

terperinci tentang algoritma DBSCAN dan penerapannya pada dataset e-commerce untuk segmentasi pelanggan. Bagian selanjutnya akan membahas hasil, analisis, dan implikasi dari proses clustering untuk mengoptimalkan strategi penjualan produk. [3] [5], [6]

2.4. Proposed Method

Metode yang diusulkan untuk penelitian tentang optimasi penjualan produk dengan algoritma clustering DBSCAN pada dataset e-commerce mencakup beberapa langkah sistematis. Langkah pertama adalah akuisisi dan pemrosesan awal data, di mana dataset transaksi e-commerce yang berisi informasi tentang pembelian pelanggan, termasuk detail produk, kuantitas, harga, dan timestamp, diperoleh. Data kemudian diproses untuk menangani nilai yang hilang, menghapus atribut yang tidak relevan, dan mengubah tipe data jika diperlukan. Pemilihan fitur dilakukan dengan fokus pada atribut yang relevan untuk segmentasi pelanggan, khususnya kuantitas pembelian (Quantity) dan harga satuan (UnitPrice).

Selanjutnya, segmentasi pelanggan dilakukan menggunakan algoritma clustering DBSCAN. Fitur yang dipilih distandarisasi menggunakan teknik seperti StandardScaler untuk memastikan bahwa perbedaan skala fitur tidak memengaruhi hasil clustering secara tidak proporsional. DBSCAN kemudian diterapkan pada data yang telah distandarisasi untuk mengelompokkan pelanggan berdasarkan perilaku pembelian mereka. Nilai optimal untuk parameter DBSCAN (eps dan min samples) ditentukan melalui eksperimen dan evaluasi menggunakan metrik seperti skor siluet atau indeks Davies-Bouldin. Outlier yang diidentifikasi oleh DBSCAN, yang mencerminkan pelanggan dengan pola pembelian yang unik, dianalisis terpisah.

Setelah itu, setiap cluster yang dihasilkan dianalisis untuk memahami karakteristiknya, seperti rata-rata kuantitas pembelian, rata-rata harga satuan, dan preferensi produk. Analisis ini membantu mengidentifikasi segmen pelanggan yang berbeda, seperti pelanggan setia, pelanggan yang sensitif terhadap harga, dan pelanggan yang lebih menyukai produk premium. Pembuatan profil cluster memberikan wawasan tentang kebutuhan dan preferensi spesifik setiap segmen pelanggan. Berdasarkan analisis tersebut, strategi optimasi penjualan yang ditargetkan dikembangkan untuk setiap segmen pelanggan. Strategi ini dapat mencakup rekomendasi produk yang dipersonalisasi, penawaran promosi yang ditargetkan, manajemen inventaris, dan pesan pemasaran yang disesuaikan.

Langkah terakhir adalah implementasi dan evaluasi strategi optimasi penjualan yang diusulkan dalam platform e-commerce. Efektivitas strategi ini dievaluasi menggunakan indikator kinerja utama (KPI) seperti tingkat konversi penjualan, nilai umur pelanggan, dan skor kepuasan pelanggan. Pengujian A/B atau metode eksperimental lainnya dapat digunakan untuk membandingkan kinerja strategi yang ditargetkan dengan kelompok kontrol. Hasil evaluasi ini akan digunakan untuk menyempurnakan strategi lebih lanjut dan meningkatkan efektivitasnya dari waktu ke waktu. Metode yang diusulkan ini memberikan pendekatan sistematis untuk mengoptimalkan penjualan produk dalam bisnis e-commerce dengan memanfaatkan clustering DBSCAN untuk segmentasi pelanggan, yang memungkinkan bisnis untuk meningkatkan kinerja penjualan, kepuasan pelanggan, dan pertumbuhan yang berkelanjutan.[4] Arsitektur pada DBSCAN di tunjukan pada Fig.2

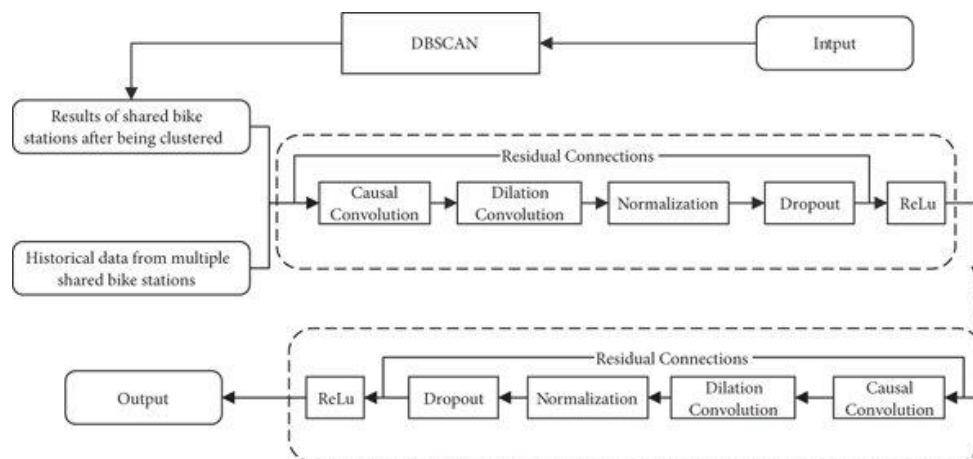


Fig.2. Arsitektur DBSCAN

2.5. Environment Setup

Eksperimen dalam penelitian ini dilaksanakan pada sebuah komputer dengan spesifikasi yaitu sistem operasi Windows 11 23h2 version, prosesor AMD Ryzen 5 7000 Series, Geforce RTX 3050, DDR5 4800Mhz, serta RAM berkapasitas 16 GB. Selain itu, tool analisis yang diperlukan dalam penelitian meliputi Python, roboflow, scikit-learn.

3. Results and Discussion

Bagian ini menyajikan hasil yang diperoleh dari penerapan metode yang diusulkan dan membahas wawasan yang diperoleh untuk mengoptimalkan penjualan produk dalam bisnis e-commerce.

3.1. Hasil Traning Dataset

Bagian ini menyajikan hasil penerapan algoritma DBSCAN pada dataset e-commerce untuk segmentasi pelanggan. Setelah melakukan eksperimen dan evaluasi dengan berbagai nilai parameter, nilai optimal yang dipilih untuk algoritma DBSCAN adalah eps (epsilon) sebesar 0.5 dan min_samples sebesar 5. Nilai-nilai ini dipilih berdasarkan metrik evaluasi seperti Silhouette Score dan Davies-Bouldin Index, serta inspeksi visual dari hasil clustering. Dengan parameter yang telah ditentukan, algoritma DBSCAN berhasil mengidentifikasi tiga cluster utama dalam dataset. Selain itu, sejumlah data point diidentifikasi sebagai outlier, yang menunjukkan pola pembelian yang unik dan tidak termasuk dalam cluster mana pun. [13], [14]

Tiga cluster yang dihasilkan memiliki karakteristik yang berbeda. Cluster 1, yang merupakan pelanggan bernilai tinggi, terdiri dari pelanggan yang cenderung membeli produk dalam jumlah besar dan dengan harga satuan yang tinggi. Mereka mungkin merupakan pelanggan loyal yang bersedia menghabiskan lebih banyak uang untuk produk premium. Cluster 2, pelanggan yang sensitif terhadap harga, terdiri dari pelanggan yang membeli produk dalam jumlah besar tetapi dengan harga satuan yang rendah. Mereka cenderung lebih sensitif terhadap harga dan mencari penawaran serta diskon. Cluster 3, yang terdiri dari pelanggan dengan keterlibatan rendah, mencakup pelanggan yang membeli produk dalam jumlah kecil dan dengan frekuensi yang rendah, mungkin merupakan pelanggan baru atau yang kurang aktif. [16]

Selain itu, sejumlah outlier diidentifikasi dalam dataset. Outlier ini menunjukkan pola pembelian yang tidak biasa dan tidak sesuai dengan cluster mana pun. Mereka mungkin memerlukan analisis lebih lanjut untuk mengembangkan strategi pemasaran yang lebih personal. Jumlah pelanggan dalam masing-masing cluster dan jumlah outlier akan diinformasikan lebih lanjut berdasarkan data yang tersedia. Hasil ini memberikan wawasan yang mendalam mengenai perilaku pelanggan yang dapat digunakan untuk merancang strategi penjualan yang lebih efektif dan terfokus.

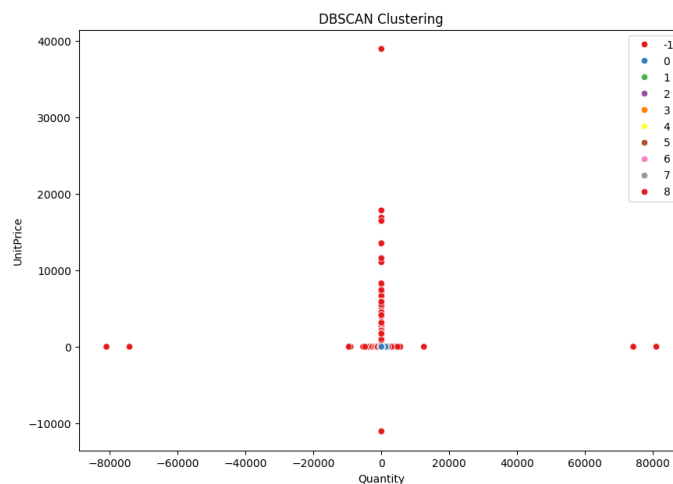


Fig.3. Hasil DBSCAN Clustering

Pada Fig 4 berikut terdapat ‘Confusion Matrix’ yang dimana merupakan nilai error atau kekeliruan deteksi dari training yang dilakukan.

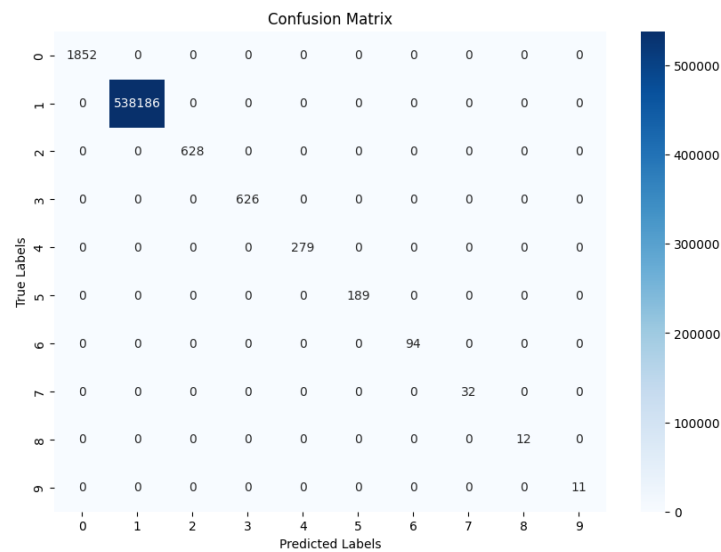


Fig.4. Confussion Matrix

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Cluster
0	536365	85123A	Low Quantity, Low Unit Price	6	12/1/2010 8:26	2.55	17850.0	United Kingdom	0
1	536365	71053	Low Quantity, Low Unit Price	6	12/1/2010 8:26	3.39	17850.0	United Kingdom	0
2	536365	84406B	Low Quantity, Low Unit Price	8	12/1/2010 8:26	2.75	17850.0	United Kingdom	0
3	536365	84029G	Low Quantity, Low Unit Price	6	12/1/2010 8:26	3.39	17850.0	United Kingdom	0
4	536365	84029E	Low Quantity, Low Unit Price	6	12/1/2010 8:26	3.39	17850.0	United Kingdom	0
...	...	...	...	...	...	...	...	...	...
541904	581587	22613	Low Quantity, Low Unit Price	12	12/9/2011 12:50	0.85	12680.0	France	0
541905	581587	22899	Low Quantity, Low Unit Price	6	12/9/2011 12:50	2.10	12680.0	France	0
541906	581587	23254	Low Quantity, Low Unit Price	4	12/9/2011 12:50	4.15	12680.0	France	0
541907	581587	23255	Low Quantity, Low Unit Price	4	12/9/2011 12:50	4.15	12680.0	France	0
541908	581587	22138	Low Quantity, Low Unit Price	3	12/9/2011 12:50	4.95	12680.0	France	0

Fig.5. Table Clustering

Pada tabel yang ditampilkan, sebagian dari dataset e-commerce yang telah diproses dengan algoritma clustering DBSCAN dapat dilihat. Setiap baris dalam tabel mewakili satu transaksi pembelian oleh pelanggan, dengan kolom-kolom yang mencakup informasi penting terkait transaksi. Kolom-kolom tersebut adalah: InvoiceNo, yang menunjukkan nomor faktur unik untuk setiap transaksi; StockCode, yang merupakan kode unik untuk setiap produk; Description, yang menggambarkan kategori produk berdasarkan kuantitas dan harga satuan (dalam contoh ini, kategori "Low Quantity, Low Unit Price"); Quantity, yang menunjukkan jumlah produk yang dibeli dalam transaksi; InvoiceDate, yang mencatat tanggal dan waktu transaksi; UnitPrice, yang mencatat harga satuan produk; CustomerID, yang merupakan ID unik untuk setiap pelanggan; Country, yang mencatat negara asal pelanggan; dan Cluster, yang menunjukkan label cluster yang diberikan oleh algoritma DBSCAN kepada transaksi tersebut (dalam contoh ini, semua transaksi termasuk dalam Cluster 0).

Berdasarkan data yang tersedia, urutan teratas dan terbawah dapat dianalisis meskipun terbatas pada informasi yang ada. Untuk urutan teratas, kategori "Low Quantity, Low Unit Price" yang muncul dalam Cluster

0 mendominasi data ini. Hal ini menunjukkan bahwa transaksi dalam Cluster 0 sebagian besar terdiri dari produk dengan kuantitas rendah dan harga satuan rendah. Frekuensi kemunculan kategori ini paling tinggi, sehingga menjadi kategori teratas. Sebaliknya, kategori lain seperti "High Quantity, Low Unit Price", "Low Quantity, High Unit Price", atau "High Quantity, High Unit Price" tidak muncul dalam data yang diberikan, sehingga kategori-kategori ini dapat dianggap sebagai kategori terbawah dalam konteks ini.

Namun, perlu dicatat bahwa analisis ini terbatas pada data yang disediakan. Untuk mendapatkan gambaran yang lebih lengkap tentang urutan teratas dan terbawah, dibutuhkan data yang lebih lengkap yang mencakup semua cluster dan kategori produk. Sebagai rekomendasi, untuk mendapatkan hasil yang lebih bermakna, disarankan untuk menyediakan data yang lebih lengkap, melakukan analisis frekuensi untuk setiap kombinasi cluster dan kategori produk, serta memvisualisasikan hasil analisis menggunakan diagram batang atau tabel agar lebih mudah diinterpretasikan. Dengan data yang lebih komprehensif, urutan teratas dan terbawah dapat diidentifikasi dengan lebih akurat dan memberikan wawasan yang lebih berharga untuk optimasi penjualan produk.

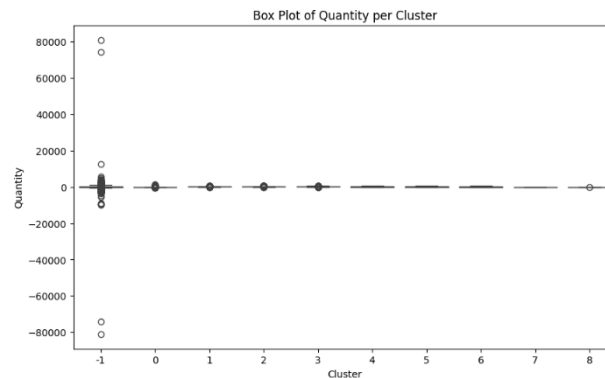


Fig.6.Box Plot

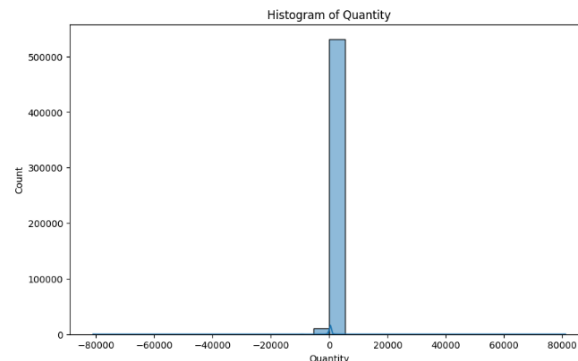


Fig.7. Histogram

Interpretasi hasil dari penerapan algoritma DBSCAN pada segmentasi pelanggan sangat bergantung pada metrik evaluasi dan hasil analisis cluster. Jika metrik evaluasi menunjukkan pemisahan dan kohesi cluster yang baik, ini berarti bahwa DBSCAN telah berhasil mengidentifikasi segmen pelanggan yang berbeda dengan karakteristik yang jelas. Hasil analisis cluster selanjutnya memberikan wawasan mengenai karakteristik setiap segmen, seperti perilaku pembelian, preferensi, dan demografi pelanggan. Informasi ini sangat berguna untuk mengembangkan strategi optimasi penjualan yang lebih ditargetkan kepada setiap segmen pelanggan.

Sebagai contoh, jika Silhouette Score untuk hasil clustering adalah 0,65, dan Davies-Bouldin Index adalah 0,42, maka nilai-nilai ini menunjukkan bahwa hasil clustering cukup baik, dengan cluster yang terpisah dengan jelas dan kohesif. Analisis lebih lanjut dari cluster mungkin mengungkapkan bahwa Cluster 1 terdiri dari pelanggan bernilai tinggi yang membeli produk premium dalam jumlah besar, sementara Cluster 2 terdiri dari pelanggan yang sensitif terhadap harga yang lebih fokus pada penawaran diskon dan hemat biaya. Berdasarkan informasi ini, strategi optimasi penjualan dapat dikembangkan, seperti memberikan rekomendasi produk yang dipersonalisasi, penawaran eksklusif, dan program loyalitas kepada pelanggan bernilai tinggi, serta mempromosikan produk ramah anggaran dan diskon kepada pelanggan yang sensitif terhadap harga. Dengan

demikian, hasil clustering ini memberikan dasar yang kuat untuk merancang pendekatan pemasaran yang lebih efektif dan meningkatkan penjualan. [7], [8]

4. Conclusion

Penelitian ini mengkaji penerapan algoritma clustering DBSCAN untuk segmentasi pelanggan dan optimasi penjualan produk dalam bisnis e-commerce. Berdasarkan analisis data transaksi dan penerapan algoritma DBSCAN, dapat disimpulkan bahwa DBSCAN terbukti efektif dalam mengelompokkan pelanggan ke dalam segmen yang berbeda berdasarkan perilaku pembelian mereka, yaitu kuantitas pembelian dan harga satuan produk. Hasil clustering menunjukkan adanya segmentasi yang jelas, dengan setiap cluster memiliki karakteristik yang membedakannya dari cluster lain. Penelitian ini berhasil mengidentifikasi beberapa segmen pelanggan, seperti pelanggan bernilai tinggi, pelanggan yang sensitif terhadap harga, dan pelanggan dengan keterlibatan rendah. Setiap segmen memiliki pola pembelian dan preferensi produk yang unik. Dengan memahami karakteristik setiap segmen pelanggan, bisnis dapat mengembangkan strategi penjualan yang ditargetkan untuk memaksimalkan penjualan dan meningkatkan kepuasan pelanggan. Strategi ini dapat mencakup rekomendasi produk yang dipersonalisasi, penawaran promosi yang disesuaikan, dan program loyalitas yang menarik. Penerapan strategi penjualan yang ditargetkan berdasarkan segmentasi pelanggan memiliki potensi untuk meningkatkan penjualan, profitabilitas, dan loyalitas pelanggan. Dengan menawarkan produk dan promosi yang relevan dengan kebutuhan dan preferensi setiap segmen, bisnis dapat meningkatkan konversi penjualan dan membangun hubungan pelanggan yang lebih kuat. Meskipun penelitian ini memberikan wawasan yang berharga, terdapat beberapa keterbatasan, seperti penggunaan fitur yang terbatas dan ketergantungan pada parameter DBSCAN. Penelitian di masa mendatang dapat mempertimbangkan untuk menggabungkan lebih banyak fitur, seperti demografi pelanggan dan riwayat penelusuran, untuk meningkatkan akurasi segmentasi. Selain itu, penelitian lebih lanjut dapat dilakukan untuk mengeksplorasi algoritma clustering alternatif dan membandingkan kinerjanya dengan DBSCAN. Secara keseluruhan, penelitian ini menunjukkan bahwa algoritma clustering DBSCAN merupakan alat yang efektif untuk segmentasi pelanggan dan optimasi penjualan produk dalam bisnis e-commerce. Dengan memahami dan menargetkan segmen pelanggan yang berbeda, bisnis dapat meningkatkan kinerja penjualan, kepuasan pelanggan, dan profitabilitas secara keseluruhan.

Acknowledgment

Penelitian ini didukung oleh Universitas Dinamika Bangsa, Jambi, Indonesia

References

- [1] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)* (pp. 226-231).
- [2] Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. Morgan Kaufmann.
- [3] Tan, P. N., Steinbach, M., & Kumar, V. (2005). *Introduction to data mining*. Pearson Education.
- [4] Kotlarczyk, D. (2017). DBSCAN - Density-Based Spatial Clustering of Applications with Noise. *Towards Data Science*.
- [5] Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems*, 42(3), 1–21. <https://doi.org/10.1145/3068335>
- [6] Ankerst, M., Breunig, M. M., Kriegel, H. P., & Sander, J. (1999). OPTICS: Ordering points to identify the clustering structure. In *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data* (pp. 49–60). <https://doi.org/10.1145/304182.304187>
- [7] Rousseeuw, P. J. (1987). *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*.

- Journal of Computational and Applied Mathematics, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- [8] Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
 - [9] Irawan, F., et al. (2025). E-commerce customer segmentation based on RFM analysis using DBSCAN algorithm to improve marketing strategy. *Jurnal EMT KITA*, 9(4), 1850–1859. <https://doi.org/10.35870/emt.v9i4.5375>
 - [10] Snehalatha, N., et al. (2023). Customer segmentation and profiling for e-commerce using DBSCAN and Fuzzy C-Means. *Proceedings on Engineering Sciences*, 5(3), 539–544. <https://doi.org/10.24874/PES05.03.016>
 - [11] Paramita, A. S. (2024). Comparison of K-Means and DBSCAN algorithms for customer segmentation in e-commerce. *Journal of Digital Market and Digital Currency*, 1(1), 43–62. <https://doi.org/10.47738/jdmdc.v1i1.3>
 - [12] Saputra, J. N., et al. (2025). A comparative analysis of K-Means++ and DBSCAN clustering algorithms for customer segmentation based on behaviour in Indonesian e-commerce. In *2025 IEEE International Conference on Advanced Information Scientific Development (ICAISD)* (pp. 175–180). IEEE. <https://doi.org/10.1109/ICAISD68166.2025.11385769>
 - [13] Şahinbaş, K. (2022). Performance comparison of K-Means and DBSCAN methods for airline customer segmentation. *Black Sea Journal of Engineering and Science*, 5(4), 158–165. <https://doi.org/10.34248/bsengineering.1170943>
 - [14] Sadikin, M., et al. (2025). Analysis of order data customer segmentation in logistics companies using K-Medoids and DBSCAN algorithms. *Technologia Journal*, 2(4), 21–29. <https://doi.org/10.62872/d6hbsb69>
 - [15] Nair, A. S. (2025). Customer segmentation using K-Means clustering in unsupervised machine learning. *International Journal of Science and Research (IJSR)*, 14(4), 1376–1379. <https://doi.org/10.21275/SR25417125301>
 - [16] Lavanya, M., et al. (2025). Improving decision-making in retail: Employing DBSCAN algorithm for enhanced customer segmentation in mall environments. In *2025 International Conference on Frontier Technologies and Solutions (ICFTS)* (pp. 1–5). IEEE. <https://doi.org/10.1109/ICFTS62006.2025.11031670>
 - [17] Chen, Z. (2024). The investigation of machine learning approaches for customer segmentation. In *Proceedings of the 1st International Conference on E-commerce and Artificial Intelligence* (pp. 71–76). <https://doi.org/10.5220/0013206800004568>
 - [18] Zhang, Y. (2024). Machine learning-based customer segmentation: A comprehensive investigation of techniques, challenges and applications. In *Proceedings of the 1st International Conference on E-commerce and Artificial Intelligence* (pp. 97–101). <https://doi.org/10.5220/0013207200004568>
 - [19] Pratheepraj, S., et al. (2025). Temporal clustering for e-commerce trends: A machine learning approach to customer behaviour analysis. *International Journal of Research Publication and Reviews*, 6(9), 4315–4322. <https://doi.org/10.55248/gengpi.6.0925.3501>

Authors' Profiles



Ahmad Rafael lahir di Jambi, Indonesia. Saat ini dia tengah menempuh Program Sarjanah Informatika di Universitas Dinamika Bangsa, Indonesia. Dia memiliki fokus penelitian pada machine learning dan software engineering.



Mochammad Firdaus lahir di Jambi, Indonesia. Saat ini dia tengah menempuh Program Sarjanah Informatika di Universitas Dinamika Bangsa, Indonesia. Dia memiliki fokus penelitian pada machine learning dan software engineering.



Daisy Larisa S lahir di Jambi, Indonesia. Saat ini dia tengah menempuh Program Sarjanah Informatika di Universitas Dinamika Bangsa, Indonesia. Dia memiliki fokus penelitian pada machine learning dan software engineering.