

SEGMENTASI PELANGGAN MENGGUNAKAN METODE DBSCAN UNTUK MENDETEKSI POLA BELANJA

Riska Dwi Handayani¹⁾, Dwi Astuti²⁾, Wahyu Priyoatmoko³⁾, Kapti⁴⁾

¹⁾ "Manajemen Informatika" STMIK BINA PATRIA Magelang

²⁾ "Sistem Informasi" STMIK BINA PATRIA Magelang

³⁾ "Teknik Informatika" STMIK BINA PATRIA Magelang

⁴⁾ "Teknik Informatika" STMIK BINA PATRIA Magelang

Email : riska@stmikbinapatria.ac.id¹⁾, dwi.astuti@stmikbinapatria.ac.id²⁾,
wepe817@stmikbinapatria.ac.id³⁾, tensmart18@stmikbinapatria.ac.id⁴⁾

Abstract

Customer segmentation is one approach used to identify customer characteristics. Accurate segmentation allows companies to personalize offers, increase customer retention and optimize marketing costs. The purpose of this study is to group customer characteristics of a retail company using the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) method. The DBSCAN method does not require initial determination of the number of clusters, is able to recognize clusters with irregular shapes and can identify outliers or customers with extreme patterns. The dataset used is an external dataset obtained from Kaggle. The dataset contains customer personalization analysis with a total of 2,240 rows and 29 columns. The results of the study show that the DBSCAN method can produce an eps value of 1.2 and produce the highest Silhouette Score of 0.080 with 4 clusters formed. Visualization of segmentation results with PCA dimension reduction techniques into two dimensions to facilitate interpretation. The PCA visualization produces 5 clusters, each of which represents its respective customer group. Thus, this approach offers an adaptive segmentation alternative that is more sensitive to complex behavioral patterns.

Keywords : *Customer segmentation, DBSCAN, shopping patterns, silhouette score, PCA*

Abstrak

Segmentasi pelanggan merupakan salah satu pendekatan yang digunakan untuk mengenali karakteristik pelanggan. Segmentasi yang akurat memungkinkan perusahaan untuk mempersonalisasi penawaran, meningkatkan retensi pelanggan serta mengoptimalkan biaya pemasaran. Tujuan dari penelitian ini adalah mengelompokkan karakteristik pelanggan dari suatu perusahaan ritel dengan menggunakan metode DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*). Metode DBSCAN tidak mengharuskan penentuan jumlah kluster diawal, mampu mengenali kluster dengan bentuk tidak beraturan serta dapat mengidentifikasi outlier atau pelanggan dengan pola ekstrem. Dataset yang digunakan merupakan dataset eksternal yang didapat dari Kaggle. Dataset tersebut berisi analisis personalisasi pelanggan dengan jumlah 2.240 baris dan 29 kolom. Hasil penelitian menunjukkan bahwa metode DBSCAN dapat menghasilkan nilai eps sebesar 1.2 dan menghasilkan Silhouette Score tertinggi yaitu 0.080 dengan 4 cluster terbentuk. Visualisasi hasil segmentasi dengan teknik reduksi dimensi PCA kedalam dua dimensi untuk memudahkan interpretasi. Dari visualisasi PCA tersebut menghasilkan 5 buah kluster yang masing-masing kluster mewakili kelompok pelanggan masing-masing. Dengan demikian, pendekatan ini menawarkan alternatif segmentasi yang adaptif dan lebih sensitive terhadap pola perilaku yang kompleks.

Kata Kunci : *Segmentasi pelanggan, DBSCAN, pola belanja, silhouette score, PCA*

1. Pendahuluan

Dalam era persaingan bisnis yang semakin ketat, perusahaan ritel dituntut untuk memahami perilaku pelanggan secara lebih mendalam untuk menciptakan strategi pemasaran yang efektif. Salah satu pendekatan yang banyak digunakan untuk memahami karakteristik pelanggan adalah melalui segmentasi pelanggan, proses pengelompokan konsumen ke dalam kelompok-kelompok homogen berdasarkan kesamaan karakteristik tertentu seperti perilaku belanja, preferensi, produksi atau frekuensi pembelian (Kotler & Keller, 2016). Dengan memantau tren pembelian konsumen, perusahaan dapat mengenali pola penjualan. Analisis dan pengelolaan yang tepat dari data tersebut memungkinkan indentifikasi produk unggulan serta produk yang kurang diminati. Hal ini berperan penting dalam pengendalian stok dan menjadi bahan pertimbangan untuk menyusun strategi pemasaran yang efektif (Normah et al., 2020).

Segmentasi yang akurat memungkinkan perusahaan untuk mempersonalisasi penawaran, meningkatkan retensi pelanggan, serta mengoptimalkan biaya pemasaran. Namun tantangan utama dalam segmentasi adalah bagaimana mengelompokkan pelanggan yang memiliki pola belanja yang kompleks dan tidak selalu mengikuti distribusi linier atau simetris. Untuk mengatasi tantangan tersebut, metode DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) menjadi salah satu alternatif yang menjanjikan. Berbeda dengan metode klasik seperti K-Means, DBSCAN tidak mengharuskan penentuan jumlah kluster diawal, mampu mengenali kluster dengan bentuk tidak beraturan, serta dapat mengidentifikasi outlier atau pelanggan dengan perilaku ekstrem (Ester et al., 1996). Hal ini sangat relevan dalam konteks data pelanggan ritel yang cenderung memiliki variasi tinggi dalam frekuensi dan jenis pembelian.

Selain itu, DBSCAN dinilai efektif dalam mendeteksi pola belanja pelanggan berdasarkan kepadatan data, sehingga mampu memberikan *insight* yang lebih akurat tentang kelompok pelanggan aktif, pasif, hingga pelanggan bernilai tinggi (Han et al., n.d.). Dengan menerapkan metode ini, perusahaan dapat mengidentifikasi strategi pendekatan yang berbeda untuk setiap segmen, seperti penawaran khusus untuk pelanggan loyal atau program reaktivasi untuk pelanggan pasif.

Hasil pengelompokan dari algoritma DBSCAN ini selanjutnya akan dievaluasi menggunakan metode *silhouette*. *Silhouette Score* adalah metode evaluasi pengelompokan yang mengintegrasikan pendekatan kohesi dan pemisahan. Kohesi menilai seberapa rapat objek-objek dalam satu kluster, sementara pemisahan mengukur jarak rata-rata antar objek dari kluster berbeda. Setiap kluster dievaluasi menggunakan *silhouette* (Paembonan et al., 2021). Nilai *silhouette*, yang dilambangkan sebagai $s(k)$, diperoleh dari rata-rata nilai *silhouette* setiap objek dalam suatu kluster. Skor *silhouette* berada dalam rentang -1 hingga 1, dimana nilai yang mendekati 1 menunjukkan kualitas pengelompokan yang sangat baik, sementara nilai mendekati -1 mencerminkan hasil klusterisasi yang kurang optimah atau buruk (Hidayati et al., 2021).

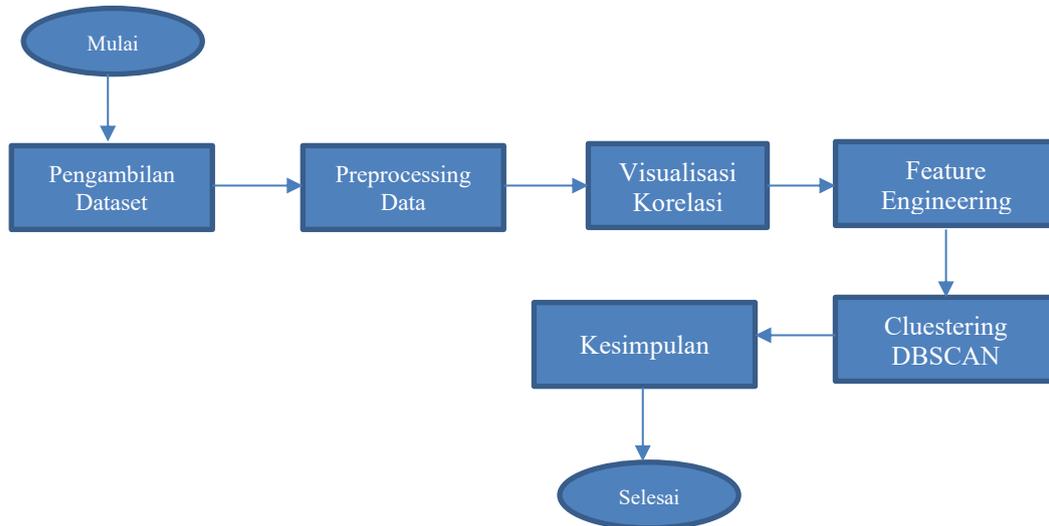
Penelitian sebelumnya yang dilakukan oleh Fajar Supriadi, menunjukkan bahwa metode DBSCAN bekerja dengan menumbuhkan area-area dengan kepadatan yang cukup tinggi ke dalam cluster-cluster dan menemukan kluster-kluster dalam bentuk sembarangan dalam suatu database spasial yang memuat noise (Supriadi, 2025). Penelitian juga dilakukan oleh (Brawijaya et al., 2017) metode DBSCAN menghasilkan clustering yang baik dengan *silhouette score* 0.767 dan nilai indeks DBI 0.731.

Berdasarkan latar belakang diatas, penelitian ini bertujuan untuk menerapkan metode DBSCAN dalam segmentasi pelanggan guna mendeteksi pola belanja yang tersembunyi dalam data, serta memberikan rekomendasi strategis berdasarkan hasil klusterisasi yang diperoleh.

2. Metode Penelitian

Pendekatan penelitian pada studi kasus ini adalah kuantitatif. Pendekatan kuantitatif digunakan untuk mengumpulkan dan menganalisis data numerik dalam rangka memahami

hasil pengelompokan menggunakan algoritma DBSCAN pada segmentasi pelanggan guna mengidentifikasi pola belanja. Melalui metode ini, data berupa angka dan statistic dijadikan dasar untuk memperoleh pemahaman yang objektif dan mendalam terhadap efektivitas algoritma tersebut. Gambaran umum tahapan penelitian yang akan dilaksanakan disajikan menggunakan diagram alir pada gambar 1.



Gambar 1. Diagram Alir Tahapan Penelitian

Keterangan diagram alir penelitian :

1. Pengambilan Dataset
Langkah ini mencakup proses pengambilan data dari sumber eksternal, yaitu platform Kaggle yang menyediakan analisi kepribadian pelanggan. Data yang didapat merupakan data pemasaran dari sebuah perusahaan ritel dan mencakup informasi demografis, kebiasaan pembeli, serta keterlibatan pelanggan terhadap berbagai saluran pemasaran.
2. *Preprocessing* Data
Langkah ini bertujuan untuk mempersiapkan data yang akan digunakan dalam proses pengelompokan (*clustering*), adapun tahapan dari preprocessing data meliputi pembersihan data kosong, menambahkan fitur umur, dan menggabungkan kolom.
3. Visualisasi Korelasi
Pada tahapan ini korelasi menggunakan Heatmap untuk mengeksplorasi hubungan antar fitur numerik dalam dataset. Visualisasi ini membantu dalam memahami sejauh mana fitur-fitur tersebut saling terkait dan dapat membantu dalam pemilihan fitur yang relevan untuk proses clustering. Dengan melihat korelasi antar fitur, dapat diidentifikasi hubungan yang kuat antar variabel dan menghindari fitur yang sangat berkorelasi (multikolinieritas) yang dapat mempengaruhi hasil model.
4. Pemilihan fitur ini didasarkan pada relevansi terhadap perilaku pelanggan, seperti pengeluaran dan cara berbelanja, serta keragaman informasi yang dapat menggambarkan karakteristik pelanggan secara lebih mendalam.
5. Clustering dengan DBSCAN
Langkah awal adalah memilih parameter Epsilon (ϵ) dan MinPts dengan menentukan nilai untuk parameter epsilon (ϵ), yang merupakan radius lingkungan, serta MinPts yang merupakan jumlah minimum titik dalam radius ϵ agar titik tersebut dapat dianggap sebagai titik inti. Kemudian, masukkan nilai parameter yang telah dipilih ke dalam model DBSCAN dan inisialisasi model dengan parameter yang telah ditentukan (Setiawan & Kurniawan,

2023). Persamaan umum yang digunakan dalam DBSCAN antara lain :

$$\epsilon - neighborhood = \{ y \in D \mid dist(x, y) \leq \epsilon \}$$

Dimana :

- ϵ : radius lingkungan
- D : dataset
- $dist(x,y)$: jarak antar x dan y

6. Evaluasi
Langkah selanjutnya adalah evaluasi dari hasil clustering.
7. Kesimpulan
Langkah terakhir adalah menyimpulkan apa dari hasil clusteringnya.

3. Hasil dan Pembahasan

3.1 Pengambilan Dataset

Dataset yang digunakan dalam penelitian ini adalah data analisis kepribadian pelanggan dari platform Kaggle. Dataset tersebut berjumlah 2.240 baris dan 29 kolom. Adapun atribut kolom ditampilkan dalam tabel 1 dibawah.

Tabel 1. Kolom Dataset

<i>Attribute</i>	<i>Keterangan</i>
<i>ID</i>	Kode unik pada setiap transaksi
<i>Year_Birth</i>	Tahun lahir pada setiap pelanggan
<i>Education</i>	Pendidikan terakhir dari pelanggan
<i>Marital_Status</i>	Status perkawinan dari pelanggan
<i>Income</i>	Pendapatan dari pelanggan
<i>Kidhome</i>	Jumlah anak kecil pelanggan
<i>Teenhome</i>	Jumlah anak remaja pelanggan
<i>Dt_Customer</i>	Tanggal mulai menjadi pelanggan
<i>Recency</i>	Jumlah hari transaksi terakhir
<i>MntWines</i>	Total pembelanjaan pelanggan untuk wine
<i>MntFruits</i>	Total pembelanjaan pelanggan untuk buah
<i>MntMeatProducts</i>	Total pembelanjaan pelanggan untuk produk daging
<i>MntFishProducts</i>	Total pembelanjaan pelanggan untuk produk ikan
<i>MntSweetProducts</i>	Total pembelanjaan pelanggan untuk produk manis
<i>MntGoldProds</i>	Total pembelanjaan pelanggan untuk produk emas
<i>NumDealsPurchases</i>	Total pembelian produk saat ada diskon
<i>NumWebPurchases</i>	Total pembelian lewat website perusahaan
<i>NumCatalogPurchases</i>	Total pembelian melalui katalog
<i>NumStorePurchases</i>	Total pembelian melalui toko fisik
<i>NumWebVisitsMonth</i>	Jumlah kunjungan pelanggan ke web perusahaan
<i>AcceptedCmp3</i>	Respon terhadap kampanye 3
<i>AcceptedCmp4</i>	Respon terhadap kampanye 4
<i>AcceptedCmp5</i>	Respon terhadap kampanye 5
<i>AcceptedCmp1</i>	Respon terhadap kampanye 1
<i>AcceptedCmp2</i>	Respon terhadap kampanye 2
<i>Complain</i>	Jumlah pengajuan complain terhdp perusahaan
<i>Z_CostContact</i>	Biaya kontak pelanggan
<i>Z_Revenue</i>	Total pendapatan yang dihasilkan dr pelanggan

<i>Response</i>	Respon terhadap kampanye terakhir
-----------------	-----------------------------------

3.2 Preprocessing Data

Tahapan preprosesing data pada penelitian ini adalah sebagai berikut :

- a. Eksplorasi Awal Dataset
 Pada eksplorasi awal ini ditemukan beberapa temuan diantaranya adalah terdapat kolom numerik, kategorikal dan tanggal yang perlu dikonversi. Kemudian pada kolom *income* memiliki 24 data kosong dan nilai maksimum yang sangat tinggi yaitu 666.666, hal ini mengindikasikan bahwa terdapat outlier pada dataset ini. Selanjutnya ditemukan juga pengeluaran tertinggi ditemukan pada produk wine dan daging.
- b. Menghapus Data Kosong
 Setelah dilakukan eksplorasi data dan ditemukan terdapat data kosong maka langkah selanjutnya adalah menghapus data yang kosong tersebut. Berikut adalah hasil pengecekan data yang hilang sebelum di hapuskan. Dari gambar 2 tersebut terlihat pada *income* terdapat kolom yang hilang sejumlah 24. Kemudian gambar 3 merupakan hasil output setelah dilakukan pembersihan sehingga terlihat sudah tidak ada kolom yang mempunyai data kosong. Hal tersebut ditandai dengan semu hasilnya adalah 0.

```
Missing value sebelum :
ID 0
Year_Birth 0
Education 0
Marital_Status 0
Income 24
Kidhome 0
Teenhome 0
Dt_Customer 0
Recency 0
MntWines 0
MntFruits 0
MntMeatProducts 0
MntFishProducts 0
MntSweetProducts 0
MntGoldProds 0
NumDealsPurchases 0
NumWebPurchases 0
NumCatalogPurchases 0
NumStorePurchases 0
NumWebVisitsMonth 0
AcceptedCmp3 0
AcceptedCmp4 0
AcceptedCmp5 0
AcceptedCmp1 0
AcceptedCmp2 0
Complain 0
Z_CostContact 0
Z_Revenue 0
Response 0
dtype: int64
```

Gambar 2. Data kosong sebelum dibersihkan

```
Missing value setelah :
ID 0
Year_Birth 0
Education 0
Marital_Status 0
Income 0
Kidhome 0
Teenhome 0
Dt_Customer 0
Recency 0
MntWines 0
MntFruits 0
MntMeatProducts 0
MntFishProducts 0
MntSweetProducts 0
MntGoldProds 0
NumDealsPurchases 0
NumWebPurchases 0
NumCatalogPurchases 0
NumStorePurchases 0
NumWebVisitsMonth 0
AcceptedCmp3 0
AcceptedCmp4 0
AcceptedCmp5 0
AcceptedCmp1 0
AcceptedCmp2 0
Complain 0
Z_CostContact 0
Z_Revenue 0
Response 0
dtype: int64
```

Gambar 3. Data Kosong setelah dibersihkan

- c. Menambahkan Fitur Umur
 Penambahna fitur umur ini dilakukan karena dalam dataset belum terdapat fitur tersebut. Penambahan fitur umur dilakukan dengan cara mengurangi tahun yang

sedang berjalan dengan Year_Birth pada dataset.

d. Menggabungkan

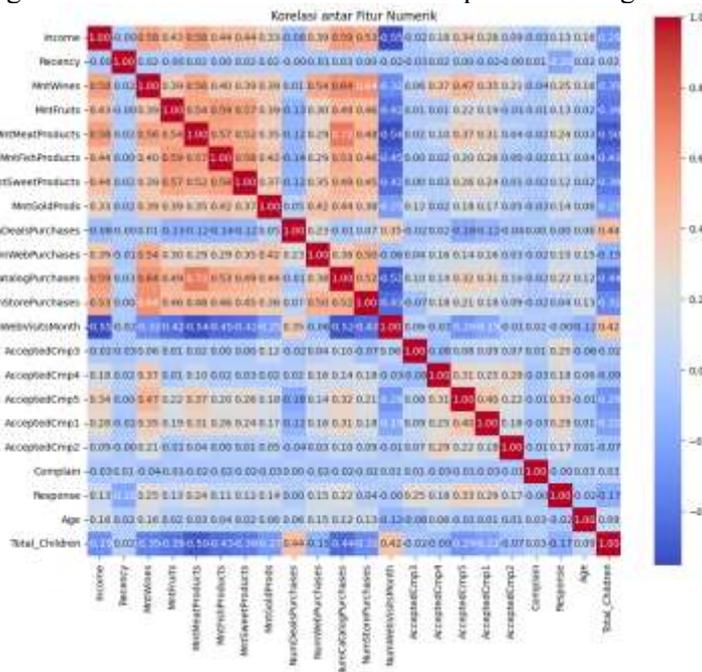
Kolom

Pada penelitian ini penggabungan kolom diperlukan karena pada dataset terdapat kolom *Kidhome* dan kolom *Teenhome* dipisah. Sehingga kedua kolom tersebut perlu digabungkan menjadi kolom baru Bernama *TotalChildren*.

3.3 Visualisasi

Korelasi

Visualisasi korelasi ini digunakan untuk mengeksplorasi hubungan antar fitur numerik dalam dataset. Visualisasi ini membantu dalam memahami sejauh mana fitur-fitur tersebut saling terkait dan dapat membantu dalam pemilihan fitur yang relevan untuk proses clustering. Hasil korelasi antar fitur numerik dapat dilihat di gambar 4.



Gambar 4. Gambar Heatmap Korelasi Numerik

Heatmap menunjukkan hubungan antar fitur seperti *MntMeatProduct*, *Income*, dan *MntWines*. Dalam gambar diatas kolom yang memiliki warna merah menandakan korelasi yang kuat yaitu mendekati 1. Sedangkan warna biru tua menandakan relasi yang negative kuat (mendekati -1). Dan yang terakhir putih menandakan tidak adanya korelasi. Dalam gambar heatmap diatas dapat dilihat terdapat hubungan yang positif antara *MntWines* (0.58), *MntFruits* (0.43), *MntMeatProducts* (0.58), *MntFishProducts* (0.44), *MntSweetProducts* (0.44) dengan *Income*. Artinya semakin tinggi pendapatan yang diterima oleh pelanggan maka semakin besar daya beli pelanggan terhadap produk-produk tersebut. Selain itu terdapat juga hubungan negative antara *Income* dengan *NumbVisitMonth* (-0.55), yang artinya semakin rendah penghasilan pelanggan justru semakin sering dalam mengunjungi website. Selanjutnya terdapat hubungan yang kuat juga dari *MntWine* dengan *MntMeatProduct* (0.56), *NumCatalogPurchase* (0.64) dan *NumStorePurchase* (0.64). Hal tersebut berarti pelanggan yang membeli *wine* biasanya juga membeli produk daging dan sering membeli lewat katalog maupun toko fisik. Kemudian *Total_Children* mempunyai hubungan yang negative dengan *Income* (-0.29) serta *NumCatalogPurchase* (-0.44). Hal ini dapat diartikan bahwa semakin banyak anak, cenderung penghasilan lebih rendah dan lebih jarang belanja melalui katalog.

3.4 Feature Engineering

Feature Engineering merupakan proses mengubah data mentah menjadi format yang lebih ringkas dan akurat untuk pemilihan fitur yang akan digunakan dalam model *machine learning* (Saputra & Yusuf, 2024). Adapun kolom yang digunakan untuk fitur input seperti gambar 5.

index	Income	Age	Total_Children	MntWines	NumCatalogPurchases	Recency
0	58138.0	68	0	635	10	58
1	46344.0	71	2	11	1	38
2	71613.0	60	0	426	2	26
3	26646.0	41	1	11	0	26
4	58293.0	44	1	173	3	94

Show 25 per page

Gambar 5. Kolom fitur input

3.5 Penerapan

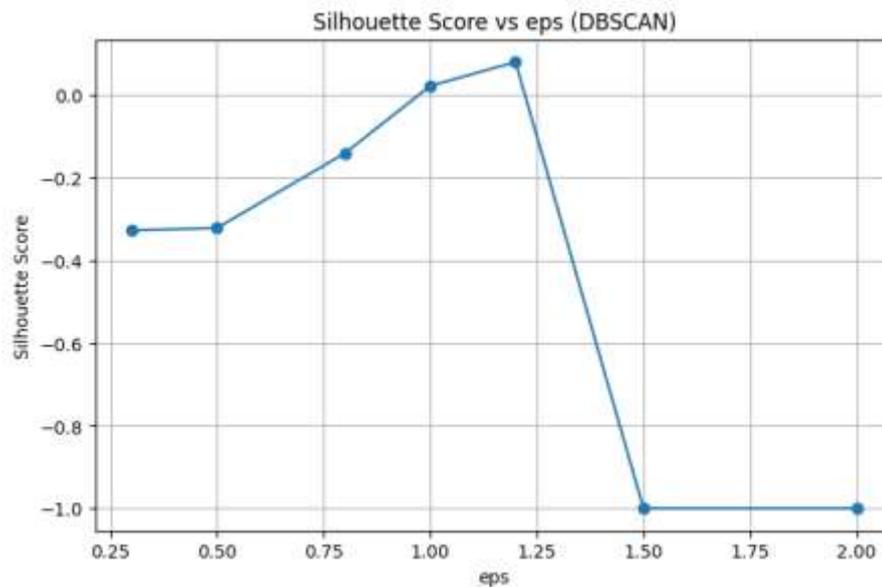
DBSCAN

Pada penelitian ini metode clustering yang digunakan adalah algoritma DBSCAN. Langkah clustering menggunakan DBSCAN dilakukan dengan 2 tahap, yaitu

a. Silhouette

Score

DBSCAN diterapkan dengan parameter eps terbaik berdasarkan nilai silhouette score tertinggi, yang menghasilkan lebih dari satu cluster. Penilaian nilai eps pada DBSCAN dapat dilihat pada gambar 6.



Gambar 6. Pemilihan nilai eps pada DBSCAN

Kemudian hasil dari pemilihan nilai eps pada DBSCAN selanjutnya difilter menggunakan data skala yang tertinggi dari data frame tersebut dan menghasilkan nilai sesuai pada gambar 7 dibawah ini

```

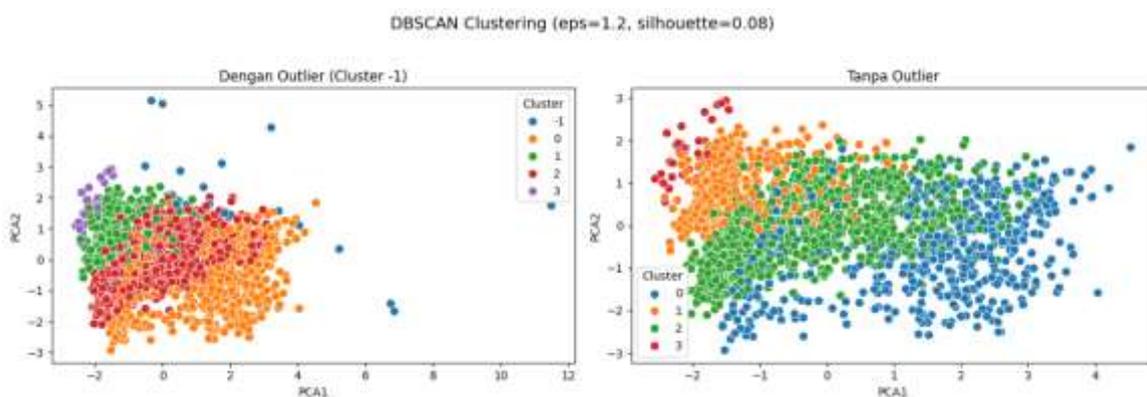
Evaluasi Clustering:
Metode      Jumlah Cluster  Silhouette Score
DBSCAN      4               0.086
    
```

Gambar 7. Hasil filter nilai eps

Berdasarkan hasil filter dan grafik diatas, nilai eps 1.2 menghasilkan *Silhouette Score* tertinggi dengan nilai 0.080 dengan 4 cluster terbentuk. Nilai eps yang terlalu kecil atau besar cenderung menghasilkan pemisahan cluster yang kurang optimah atau hanya membentuk satu klaster.

b. PCA (Principal Component Analysis)

Selanjutnya hasil clustering divisualisasikan menggunakan PCA untuk mereduksi dimensi data ke dalam dua komponen utama, sehingga memudahkan visualisasi cluster. Setiap warna pada plot merepresentasikan cluster yang terbentuk dengan outlier ditandai dengan label -1. Gambar visualisasi menggunakan PCA dapat dilihat pada gambar 7.



Gambar 8 Hasil Visualisasi PCA

Gambar diatas merupakan hasil visualisasi epsilon 1.2 yang dibuat ke dalam dua dimensi untuk memudahkan interpretasi. Setiap titik pada grafik pada grafik mewakili satu pelanggan, dan warna menandakan klaster yang terbentuk. Berikut gambaran masing-masing klaster :

- Cluster 0 (warna oranye) : Mewakili kelompok pelanggan terbesar. Umumnya terdiri dari pelanggan dengan pengeluaran moderat, aktif dalam pembelian katalog dan berpenghasilan menengah. Mereka cenderung stabil dan dapat dijadikan target utama untuk promosi umum.
- Cluster 1 (warna hijau) : pelanggan dengan jumlah anak lebih banyak dan penghasilan relatif rendah. Mereka lebih jarang melakukan pembelian dan cenderung hemat. Segmen ini dapat dijangkau melalui pendekatan promosi hemat atau diskon keluarga
- Cluster 2 (warna merah): Pelanggan dengan pengeluaran tinggi untuk produk tertentu seperti wine. Pelanggan=n juga aktif berbelanja katalog serta kemungkinan adalah pelanggan loyal dengan preferensi premium.
- Cluster 3 (warna ungu): Kelompok lebih kecil yang mungkin terdiri dari pelanggan dengan karakteristik unik, seperti lebih tua, selektif dalam pembelian, atau hanya berbelanja melalui saluran tertentu.
- Cluster -1 (outlier): Kelompok ini berisi pelanggan dengan perilaku yang tidak biasa atau ekstrem, seperti pengeluaran yang sangat tinggi atau sangat rendah, pola belanja yang tidak teratur serta jarang melakukan transaksi atau interaksi. Dalam DBSCAN cluster ini terdeteksi sebagai noise atau data yang menyimpang dari pola umum.

4. Kesimpulan

Penelitian ini bertujuan untuk mengetahui pola belanja dalam segmentasi pelanggan menggunakan algoritma DBSCAN. Berdasarkan pembahasan diatas algoritma DBSCAN dapat menghasilkan nilai eps sebesar 1.2 menghasilkan *silhouette score* tertinggi yaitu 0.080 dengan 4 cluster terbentuk. Hal ini menunjukkan bahwa pemisahan kluster cukup baik, dengan struktur kepadatan data yang mulai terbentuk secara jelas. Algoritma DBSCAN mampu mengelompokkan pola belanja tanpa perlu menentukan jumlah kluster diawal. Metode ini tidak hanya membentuk kelompok pelanggan dengan karakteristik yang serupa tetapi juga dapat mendeteksi *outlier* yaitu pelanggan dengan karakteristik yang berbeda atau karakteristik minoritas. Selanjutnya informasi hasil klasterisasi ini dapat dimanfaatkan oleh perusahaan untuk Menyusun strategi promosi yang lebih tepat sasaran, mempertahankan pelanggan setia, serta mengenali pelanggan dengan karakteristik minoritas.

Daftar Pustaka

- Brawijaya, U., Maulana Pranata, F., Hadi Wijoyo, S., & Setiawan, N. Y. (2017). Fakultas Ilmu Komputer Analisis Performa Algoritma K-Means dan DBSCAN Dalam Segmentasi Pelanggan Dengan Pendekatan Model RFM (Vol. 1, Issue 1). <http://j-ptiik.ub.ac.id>
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. www.aaai.org
- Han, J., Kamber, M., & Pei, J. (n.d.). *Data Mining Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Hidayati, R., Zubair, A., Hidayat Pratama, A., Indana, L., Studi Sistem Informasi, P., & Teknologi Informasi, F. (2021). Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering Silhouette Coefficient Analysis in 6 Measuring Distances of K-Means Clustering (Vol. 20, Issue 2).
- Kotler, P., & Keller, K. L. (2016). *Marketing Management* (15th ed.). Pearson Education.
- Normah, Rifai, B., & Sari, P. (2020). Algoritma Apriori Sebagai Solusi Kontrol Persediaan Suku Cadang Mobil PT. Buanasakti Aneka Motor Jakarta. 22.
- Paembonan, S., Abduh, H., & Kunci, K. (2021). Penerapan Metode Silhouette Coefficient Untuk Evaluasi Clustering Obat Clustering; K-means; Silhouette coefficient (Vol. 6, Issue 2). <https://ojs.unanda.ac.id/index.php/jiit/index>
- Saputra, A., & Yusuf, R. (2024). Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1346–1361. <https://doi.org/10.57152/malcom.v4i4.1516>
- Setiawan, K. E., & Kurniawan, A. (2023). Jurnal Informatika Terpadu Pengelompokan Rumah Sakit Di Jakarta Menggunakan Model Dbscan, Gaussian Mixture, Dan Hierarchical Clustering. *Jurnal Informatika Terpadu*, 9(2), 149–156. <https://journal.nurulfikri.ac.id/index.php/JIT>
- Supriadi, fajar. (2025). Penerapan Metode Dbscan Dalam Penentuan Mahasiswa Yang Layak Memperoleh Bidikmisi20250421.