

PENERAPAN ALGORITMA DBSCAN DALAM PENGELOMPOKAN NEGARA MITRA DAGANG STRATEGIS INDONESIA

Sujito¹⁾, Nensy Enlanda Putri Pratama²⁾, Sigit Setyowibowo³⁾

^{1, 3)} *Teknologi Informasi STMIK PPKIA Pradnya Paramita*

²⁾ *Sistem Informasi STMIK PPKIA Pradnya Paramita*

email : sujito@stimata.ac.id¹⁾, nensy@stimata.ac.id²⁾, sigit@stimata.ac.id³⁾

Abstraksi

Penelitian ini bertujuan mengidentifikasi pola dan karakteristik pengelompokan negara mitra dagang strategis Indonesia berdasarkan data ekspor non-migas periode 2019–2023. Menggunakan data sekunder dari Kementerian Perdagangan, penelitian ini melakukan Algoritma DBSCAN yang berbasis kepadatan (*density-based*). Proses penentuan jumlah *cluster* optimal menggunakan *Elbow Method*, evaluasi performa metode diukur melalui *Silhouette Coefficient*. Hasil penelitian menunjukkan algoritma DBSCAN menghasilkan 9 *cluster* serta 1 *cluster* berlabel "-1" yang mengidentifikasi adanya titik noise atau *outlier* pada data. Hasil penelitian ini diharapkan dapat memberikan pandangan strategis bagi para pengambil kebijakan dalam perdagangan internasional. Dengan pemetaan yang akurat melalui penerapan metode ini, pemerintah dapat merumuskan strategi keputusan yang lebih tepat sasaran untuk meningkatkan nilai dan volume ekspor non-migas Indonesia di pasar global secara signifikan.

Kata Kunci :

Klastering, Ekspor Non-Migas, Koefisien Silhouette

Abstract

This study aims to identify the patterns and characteristics of Indonesia's strategic trading partner countries based on non-oil and gas export data for the 2019–2023 period. Using secondary data from the Ministry of Trade, this study uses the density-based DBSCAN algorithm. The process of determining the optimal number of clusters uses the Elbow Method, and the method's performance is evaluated using the Silhouette Coefficient. The results show that the DBSCAN algorithm produces 9 clusters and 1 cluster labeled "-1" which identifies the presence of noise or outliers in the data. The results of this study are expected to provide strategic insights for policymakers in international trade. With accurate mapping through the application of this method, the government can formulate more targeted decision strategies to significantly increase the value and volume of Indonesia's non-oil and gas exports in the global market.

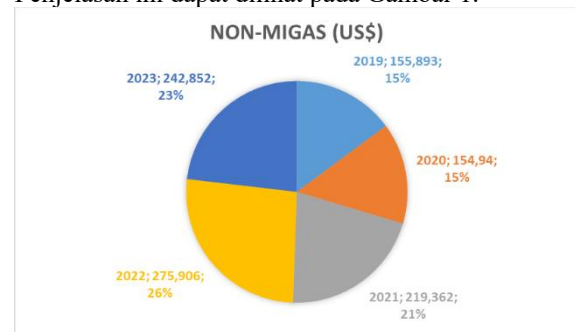
Keywords :

Clustering, Non-Oil And Gas Export, Silhouette Coefficient.

Pendahuluan

Indonesia adalah negara dengan jumlah penduduk yang besar dan pertumbuhan ekonomi yang cukup pesat [1]. Selain itu, Indonesia termasuk dalam wilayah Asia Tenggara dan dikenal sebagai perekonomian yang memiliki variasi dan sumber daya alam yang sangat kaya [2]. Pada data KEMENDAG RI, ekspor non-migas meliputi sektor industri, pertanian, dan pertambangan. Kegiatan ini tidak hanya bertujuan untuk meningkatkan devisa dan pertumbuhan ekonomi negara, tetapi juga berfungsi sebagai alat stabilitas harga dan penciptaan lapangan kerja. Di sisi lain, keberhasilan ekspor juga memperkuat citra produk lokal di mata dunia dan cinta terhadap produk lokal[3]. Ekspor di sektor non-migas memiliki potensi besar untuk perdagangan Indonesia, sehingga struktur ekspor negara ini masih didominasi oleh sektor tersebut. Berdasarkan informasi mengenai pertumbuhan ekspor non-migas antara tahun 2019 dan 2023 di KEMENDAG RI, terjadi penurunan pada tahun 2020 dari 155.893 US\$ menjadi 154.940 US\$. Namun, pada tahun 2021 hingga 2022, jumlah itu kembali naik hingga mencapai 275.906 US\$. Sementara itu, pada tahun

2023, terjadi penurunan lagi sebesar 33.053,6 US\$ dibandingkan dengan dua tahun sebelumnya. Penjelasan ini dapat dilihat pada Gambar 1.



Gambar 1. Data perkembangan ekspor non-migas tahun 2019-2023

Pada informasi tersebut, ekspor non-migas menurut negara tujuan menunjukkan bahwa data tersebut sangat krusial untuk mengidentifikasi mitra dagang utama Indonesia dan mendukung pertumbuhan ekonomi negara. Penelitian ini mengedepankan aspek perdagangan global dengan memanfaatkan data ekspor yang berasal dari KEMENDAG RI, mengacu pada perkembangan ekspor non-migas sesuai dengan negara tujuan,

mengingat bahwa Indonesia cenderung lebih banyak mengekspor bahan non-migas. Data yang diperoleh belum memberikan penjelasan mengenai pola dan karakteristik pengelompokan negara mitra dagang utama Indonesia berdasarkan tren perkembangan data ekspor non-migas, termasuk negara tujuan yang paling sering menerima ekspor dari Indonesia guna penentuan mitra dagang strategis. Oleh karena itu, penelitian ini bertujuan untuk mengelompokkan mitra dagang yang strategis bagi Indonesia dengan harapan dapat meningkatkan perekonomian dan menghindari penurunan signifikan dalam volume ekspor.

Objek penelitian ini adalah pertumbuhan ekspor non-migas (tujuan negara) selama periode 2019 hingga 2023. Data sekunder yang digunakan diambil dari informasi ekspor yang dikumpulkan oleh Kementerian Perdagangan Republik Indonesia melalui situs web berikut <https://satudata.kemendagri.go.id>. Dalam penelitian ini, metode yang diterapkan adalah pengelompokan dengan menerapkan metode DBSCAN.

Penelitian ini memfokuskan pada pengelompokan negara mitra dagang strategis Indonesia menggunakan algoritma DBSCAN, dengan bantuan bahasa pemrograman Python. Data yang digunakan merupakan data sekunder ekspor non-migas dari KEMENDAG RI dalam rentang waktu 2019–2023, yang mencakup atribut nama negara, tahun, serta jumlah nilai ekspor dalam satuan US\$, yang diarahkan untuk mengenali pola karakteristik serta potensi mitra dagang melalui tren perkembangan ekspor. Selain itu, teknis penelitian dibatasi pada penggunaan *elbow method* untuk penentuan nilai k dan metode *Silhouette Coefficient* sebagai parameter evaluasi performa model *Clustering* yang dihasilkan.

Tujuan dari penelitian ini menentukan nilai k optimal pada algoritma DBSCAN yang menghasilkan pengelompokan negara mitra dagang strategis Indonesia berdasarkan perkembangan ekspor non-migas dan mengidentifikasi pola serta karakteristik pengelompokan melalui tren data ekspor, sekaligus melakukan kinerja DBSCAN dengan penekanan pada evaluasi mendalam menggunakan nilai *Silhouette Score*.

Penelitian oleh Fadlilah (2022) [4] dalam studinya membuktikan bahwa algoritma DBSCAN efektif mengidentifikasi anomali data akademik, dengan hasil temuan outlier sebesar 5% dari total data. Penelitian tersebut menegaskan kemampuan DBSCAN dalam mendeteksi titik-titik yang berada di luar jangkauan *cluster* secara akurat.

Penelitian ini diharapkan memberikan kontribusi signifikan, baik bagi praktisi industri, pengembangan ilmu pengetahuan, maupun instansi pemerintah. Bagi perusahaan yang terlibat dalam ekspor non-migas, hasil penelitian ini dapat menjadi referensi dalam pengambilan keputusan strategis untuk menentukan negara tujuan mitra dagang. Dari sisi akademis, studi ini diharapkan dapat memperkaya literatur dan pengetahuan mengenai penerapan

algoritma DBSCAN dalam metode *clustering*. Sementara itu, bagi KEMENDAG RI, hasil pemetaan ini dapat digunakan untuk mengidentifikasi negara-negara yang memiliki potensi besar sebagai mitra dagang strategis guna mengoptimalkan kinerja ekspor non-migas Indonesia di masa depan.

Tinjauan Pustaka

Data Mining

Knowledge Discovery in Database (KDD) adalah metode untuk mengolah data yang saling berhubungan dalam tabel basis data menjadi informasi yang dapat digunakan oleh pengambil keputusan. Data mining sebenarnya adalah bagian dari kerangka kerja KDD, sehingga keduanya saling terkait meski memiliki definisi yang tidak identik [5].

Pada industri informasi menunjukkan daya tarik yang besar terhadap *data mining*, terutama dalam mengoptimalkan penyimpanan data guna menyajikan informasi yang akurat. Tujuan utama dari proses ini adalah mengidentifikasi tren atau pola dalam basis data berskala besar yang berfungsi sebagai instrumen pendukung pengambilan keputusan strategis di masa depan[6].

Berikut beberapa contoh yang termasuk dalam *data mining* atau bukan *data mining*: *Data Mining* contohnya seperti hasil laporan keuangan yang digunakan untuk prediksi pengeluaran selanjutnya dan yang bukan merupakan *data mining* seperti membandingkan laporan keuangan bulan februari dan maret untuk mencari kenaikan peminatan[7].

Clustering

Sebagai tahapan awal dalam data mining, hasil *clustering* sangat ditentukan oleh metode yang digunakan karena setiap metode memiliki kemampuan berbeda dalam mengungkap pola tersembunyi pada data. Salah satu teknik untuk mengukur tingkat kemiripan antar objek adalah *Euclidean Distance*, bekerja dengan cara menghitung jarak antara dua titik berdasarkan nilai atribut masing-masing[8]. Pengukuran eksternal dalam analisis *clustering* dilakukan dengan tolak ukur tingkat akurasi berdasarkan data di luar proses pengelompokan. Metode ini hanya dapat diterapkan pada *supervised dataset* yang memiliki label kelas. Namun, untuk menentukan jumlah kluster paling optimal tanpa ketergantungan pada label, dapat digunakan teknik *Silhouette Coefficient* dan metode *Elbow*[9].

Algoritma DBSCAN

Metode ini mengartikan kluster sebagai sekumpulan titik yang memiliki keterhubungan kepadatan (*density*) secara maksimal. Dalam penerapannya, algoritma DBSCAN bergantung pada dua parameter utama, yaitu Eps (*Epsilon*) yang menentukan radius jangkauan maksimum, serta MinPts (*Minimum Points*) yang menetapkan jumlah minimum titik dalam jangkauan tersebut. Sebuah titik dianggap dapat dijangkau secara kepadatan (*density-reachable*) apabila terdapat rangkaian titik yang saling terhubung secara langsung (*directly density-*

reachable) dari satu titik ke titik berikutnya (Fadlilah, 2022). Berikut langkah-langkah algoritma DBSCAN:

1. Menetapkan parameter utama, yaitu epsilon (ϵ) dan $minPts$.
2. Mengambil 1 sampel data secara acak untuk diperiksa.
3. Mengukur jarak antar titik menggunakan rumus Jarak *Euclidean*.
4. Menentukan kumpulan titik yang masuk dalam cakupan densitas (*density-reachable*) dari titik p .
5. Membentuk kluster baru jika titik tersebut memenuhi syarat sebagai *core point* (jumlah tetangga lebih dari sama dengan $minPts$).
6. Melewati titik jika dikategorikan sebagai *border point* yang tidak memiliki koneksi densitas, lalu melanjutkan ke data lain.
7. Melakukan pengulangan prosedur hingga seluruh populasi data terklasifikasi.

Konsep Elbow

Dalam analisis *cluster*, metode *Elbow* digunakan untuk menentukan jumlah kluster optimal dengan cara mengevaluasi persentase perbandingan hasil pada grafik *SSE* (*Sum of Squared Error*). Metode ini bekerja dengan menghitung nilai kesalahan (*error*) pada berbagai nilai k . Jumlah kluster terbaik diidentifikasi melalui titik pada grafik yang membentuk struktur menyerupai "siku" (*elbow*), di mana penurunan nilai *SSE* mulai melambat secara signifikan setelah titik tersebut [4]. Rumus *SSE* ditulis dibawah ini:

$$SSE = \sum_{k=1}^K \sum_{x_i} |x_i - c_k|^2$$

Rumus 1. *SSE* (*Sum of Square Error*)

Konsep Silhouette Coefficient

Silhouette Coefficient merupakan metrik dalam analisis data yang berfungsi untuk mengukur tingkat kecocokan suatu objek terhadap klasternya sendiri dibandingkan dengan kluster lain [10]. Dalam penelitian ini, metrik tersebut digunakan sebagai instrumen validasi internal untuk menilai kualitas hasil pengelompokan. Menurut [11], terdapat beberapa tahapan dalam melakukan evaluasi ini:

Perhitungan Nilai Silhouette:

Nilai ini dihitung untuk setiap objek menggunakan rumus berikut:

$$S_i = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Rumus 2. menghitung nilai *Silhouette*

Keterangan:

- $a(i)$: Rata-rata jarak objek i ke semua objek lain dalam kluster yang sama.
- $b(i)$: Rata-rata jarak objek i ke objek di kluster terdekat lainnya.
- Nilai *Silhouette* berada pada rentang $-1 \leq s(i) \leq 1$

Perhitungan Lebar Silhouette: Menghitung nilai rata-rata *silhouette* dari seluruh objek yang terdapat dalam setiap kluster.

Perhitungan Nilai Silhouette Coefficient

Rumus *Silhouette Coefficient* ditulis dibawah ini:

$$SC = \max s(i)$$

Rumus 3. *Silhouette Coefficient*

Skor akhir diperoleh dari nilai rata-rata terbesar di antara seluruh kluster yang terbentuk. Kualitas hasil pengelompokan kemudian ditentukan berdasarkan tabel interpretasi berikut:

Tabel 1. Interpretasi Nilai <i>Silhouette Coefficient</i>	
<i>Silhouette Coefficient</i>	Interpretasi
0,71 – 1,00	Cluster sangat kuat
0,51 – 0,70	Cluster layak/memadai
0,26 – 0,50	Cluster lemah
< 0,25	Hasil tidak menunjukkan struktur Cluster (tidak signifikan)

Ekspor Dalam Petumbuhan Ekonomi Indonesia

Ekspor merupakan sektor yang menjadi sektor yang bisa mendorong pertumbuhan ekonomi Indonesia. Dalam era global dengan persaingan yang ketat diperlukan sistem yang dapat membuat sektor ekspor lebih optimal, dengan metode memanfaatkan teknologi *data mining* yang berguna untuk menganalisis data histori dan indentifikasi peluang pasar baru yang akan dituju[12]

Penelitian oleh Nurina Sari dkk [13] yang berjudul “Penerapan Clustering DBSCAN Untuk Pertanian Padi Di Kabupaten Karawang” berfokus pada implementasi algoritma DBSCAN untuk memetakan lahan pertanian padi. Evaluasi performa dalam studi ini berhasil mengidentifikasi kluster-kluster berdasarkan perbedaan tingkat curah hujan, luas lahan yang berdampak pada volume produksi, serta intensitas dan jenis serangan hama pada masing-masing wilayah pertanian.

Penelitian oleh Adha dkk [14] dengan judul “Perbandingan Algoritma DBSCAN Dan K-Means Clustering Untuk Pengelompokan Kasus Covid-19 Di Dunia” bertujuan membandingkan performa kedua algoritma tersebut dalam mengelompokkan data pandemi global. Hasil penelitian menunjukkan

bahwa algoritma K-Means menghasilkan nilai validitas kluster yang lebih unggul dibandingkan DBSCAN. Berdasarkan pengujian tersebut, disimpulkan bahwa model paling optimal dicapai oleh algoritma K-Means pada konfigurasi jumlah kluster $k=8$.

Penelitian oleh Putri dkk [15] yang berjudul “Komparasi DBSCAN Dan K-Means Clustering Pada Pengelompokan Status Desa Di Jawa Tengah Tahun 2020” membandingkan efektivitas kedua metode tersebut dalam mengklasifikasikan status desa. Hasilnya menunjukkan bahwa algoritma DBSCAN memiliki performa yang lebih baik dibandingkan K-Means dalam memetakan desa-desa di Provinsi Jawa Tengah berdasarkan indikator IKE, IKL, dan IKS.”

Penelitian oleh Harjanto [11] melalui penelitiannya yang berjudul “Analisis Penetapan Skala Prioritas Penanganan Balita Stunting Menggunakan Metode DBSCAN Clustering” membangun sistem untuk menentukan prioritas penanganan stunting. Studi ini memanfaatkan algoritma DBSCAN untuk mengelompokkan ketersediaan fasilitas kesehatan—seperti jumlah tenaga medis dan sarana pendukung lainnya—guna memetakan skala prioritas penanganan balita secara lebih akurat.”

Penelitian oleh Armiady dkk [16] dalam penelitian berjudul “Analisis Metode DBSCAN Dalam Mendeteksi Data Outlier” berfokus pada evaluasi berbagai model DBSCAN melalui penyesuaian parameter epsilon dan MinPts. Penelitian ini bertujuan untuk membedah bagaimana perubahan parameter tersebut memengaruhi pola pembentukan kluster serta kemampuan algoritma dalam mendeteksi data pencilan (outlier).

Metode Penelitian

Pada penelitian ini data yang digunakan merupakan data sekunder yang didapat dari *website* resmi KEMENDAG RI, khususnya pada laporan perkembangan ekspor non-migas menurut negara tujuan periode 2019 hingga 2023. Data mentah tersebut mencakup berbagai atribut, seperti nama negara, tahun, tren (%) 2019-2023, perubahan (%) 2023-2022, serta kontribusi peran (%) ekspor. Namun, dalam proses *clustering*, atribut yang difokuskan hanya mencakup nama negara, tahun, dan nilai total ekspor tahunan (US\$) guna menjaga relevansi hasil pengelompokan.

Alat yang digunakan pada penelitian antara lain:

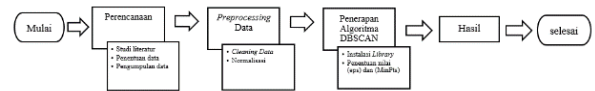
1. Laptop dengan prosesor AMD E2-2000 APU with Radeon(tm) HD Graphics 1.75 GHz dengan RAM 4 GB dan OS Windows 10 Pro
2. SSD 128GB

Sistem Pengelompokan tidak terlepas dari bahan penunjang seperti perangkat lunak yang mendukung dalam pembuatan sistem. Perangkat lunak yang digunakan antara lain:

1. Google Colab, untuk mencari nilai k dengan perhitungan *elbow* dan pengelompokan *cluster* untuk Algoritma DBSCAN.

2. Microsoft Excel 2016, untuk mengolah data (normalisasi data).

Pelaksanaan penelitian dibagi menjadi 5 tahapan yaitu: perencanaan, *preprocessing* data, *clustering*, penerapan Algoritma DBSCAN dan hasil. Tahapan penelitian disajikan dalam Gambar 2.



Gambar 2. Tahapan Penelitian

1. Tahapan Perencanaan dan Pengumpulan Data

Tahap ini diawali dengan studi literatur untuk membangun landasan teoretis yang komprehensif melalui peninjauan teori yang relevan dan hasil penelitian terdahulu. Langkah ini bertujuan untuk memetakan metodologi serta mengidentifikasi research gap atau kekurangan pada penelitian sebelumnya guna menentukan fokus penelitian saat ini. Selanjutnya, dilakukan pencarian data sekunder pada situs web resmi yang menyediakan data relevan, diikuti dengan proses pengumpulan data yang telah ditentukan untuk memastikan kesesuaian sumber dengan kebutuhan penelitian.

2. Tahapan Preprocessing Data

Tahap ini berfokus pada dua kegiatan utama, yaitu data *cleaning* dan normalisasi. Data *Cleaning*: Dilakukan untuk menjamin kualitas data dengan cara memperbaiki kesalahan tipografi, menghapus data duplikat, serta menangani data yang hilang (*missing values*) agar data menjadi konsisten. Normalisasi bertujuan untuk menyamakan skala data dari berbagai atribut yang memiliki satuan berbeda. Proses ini sangat penting agar setiap fitur memiliki bobot yang seimbang saat diproses oleh model *clustering*.

3. Tahapan Penerapan Algoritma DBSCAN

Alur kerjanya meliputi: Penentuan Parameter Optimal: Menggunakan bahasa pemrograman Python di platform Google Colab untuk mencari nilai kluster terbaik melalui metode *Elbow* untuk analisis jarak. Tahap perhitungan kinerja Algoritma DBSCAN menggunakan nilai *silhouette coefficient* rata-rata dari hasil clustering dengan memperhatikan nilai *silhouette*, yang berada disekitar 0 atau negatif menunjukkan bahwa *cluster* tidak optimal.

4. Tahapan Hasil

Sebagai tahapan setelah proses *clustering* yang akan mendapatkan hasil berupa beberapa kelompok data, yang akan diimplementasikan dan dianalisis lebih lanjut sesuai dengan tujuan penelitian.

Hasil dan Pembahasan

Hasil Penerapan Algoritma DBSCAN

Pada tahapan ini meliputi beberapa tahapan yaitu: Hasil instalasi library dan hasil nilai minPts dan epsilon (eps).

Hasil Instalasi Library

1. Instalasi algoritma DBSCAN di google colab, untuk script yang dijalankan ditampilkan pada gambar 3.

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.cluster import DBSCAN
from sklearn.preprocessing import StandardScaler
import openpyxl
from google.colab import files
import io
```

Gambar 3. Script Instalasi DBSCAN

2. Upload file untuk perhitungan DBSCAN yang sudah disesuaikan sebelumnya seperti pada gambar 4.

```
upload_files = files.upload()
for filename in upload_files.keys():
    if filename.endswith('.xlsx'):
        data = pd.read_excel(io.BytesIO(upload_files[filename]))
    else:
        print(f"Error: {filename} is not an Excel file.")
    print(data.head())
```

	2019	2020	2021	2022	2023
0 NEGARA	1.000000	1.000000	1.000000	1.000000	0.032940
1 REP.RAKYAT CINA	0.687638	0.622048	0.504854	0.444093	0.012286
2 AMERIKA SERIKAT	0.000386	0.340003	0.256655	0.366928	0.010962
3 INDIA	0.533465	0.430399	0.330665	0.365557	1.000000
4 JEPANG	0.268804	0.195657	0.168337	0.000189	0.584388
5 PILIPINA					

Gambar 4. Tampilan file yang digunakan

3. Melakukan ekstraksi sebagian data dari *DataFrame* Pandas kemudian mengubah formatnya menjadi array *NumPy*
4. Data yang telah diimpor ke *Google Colab* kemudian diproses menggunakan *StandardScaler*. Langkah ini berfungsi untuk menyamakan skala fitur (skalabilitas) agar data dapat dipahami dan diolah lebih akurat oleh algoritma *machine learning*. Hasilnya ditampilkan pada gambar 5.

```
# Scaling the data to bring all the
attributes to a comparable level
scaler = StandardScaler()
x_scaled = scaler.fit_transform(x)
# Normalizing the data so that
# the data approximately follows a Gaussian
distribution
x_normalized = normalize(x_scaled)
# Converting the numpy array into a pandas
DataFrame
x_normalized=pd.DataFrame(x_normalized)
x_scaled
```

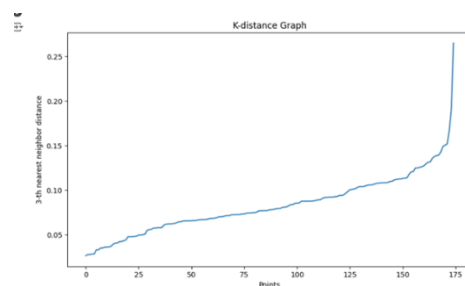
Gambar 5. Tampilan script untuk *StandardScaler*

Menentukan nilai minpts dan eps

1. Parameter epsilon ditentukan dengan mencari titik *elbow* pada *k-distance Graph* Melalui fungsi *plot_k_distance_graph*, jarak ke tetangga terdekat ke-k dihitung dan diurutkan secara sistematis. Berdasarkan kurva yang terbentuk pada grafik tersebut, diperoleh nilai epsilon sebesar 0,17 sebagai ambang batas jarak kluster yang paling sesuai. *Script* dan hasilnya ditampilkan pada gambar 6.

```
import numpy as np
from sklearn.cluster import DBSCAN
from sklearn.neighbors import NearestNeighbors
import matplotlib.pyplot as plt
# Function to plot k-distance graph
def plot_k_distance_graph(x, k):
    neigh=NearestNeighbors(n_neighbors=k)
    neigh.fit(x)
    distances, _ = neigh.kneighbors(x)
    distances = np.sort(distances[:,k-1])
    plt.figure(figsize=(10, 6))
    plt.plot(distances)
    plt.xlabel('Points')
    plt.ylabel(f'{k}-th nearest neighbor distance')
    plt.title('K-distance Graph')
    plt.show()
# Plot k-distance graph
plot_k_distance_graph(x, k=3)
```

Hasilnya



Gambar 6. Tampilan Hasil untuk penentuan epsilon

2. Nilai optimal dilakukan dengan evaluasi titik yang membentuk 'siku kurva' (*elbow of the curve*) pada grafik jarak. Setelah menetapkan nilai epsilon (eps) sebesar 0,17 pada tahap sebelumnya, langkah selanjutnya adalah menerapkan metode *Silhouette* untuk menentukan nilai *min samples* yang paling optimal. Proses ini dilakukan melalui eksekusi *script* pengujian berbagai parameter *min_samples* untuk mendapatkan kualitas *cluster* terbaik berdasarkan *silhouette score*

tertinggi dan hasilnya ditampilkan pada gambar 7.

```

from sklearn.metrics import silhouette_score
import numpy as np
# Optimal value for eps
eps = 0.17
# Range of values for min_samples to explore
min_samples_range = range(2, 10)
# Initialize a list to save silhouette scores
silhouette_scores = []
# Compute silhouette score for each
min_samples value
for min_samples in min_samples_range: DBSCAN=DB
SCAN(eps=eps,min_samples=min_samples) labels=
DBSCAN.fit_predict(x)
if len(np.unique(labels)) > 1: # Ensure
there re at least 2 clusters
silhouette_avg = silhouette_score(x, labels)
silhouette_scores.append(silhouette_avg)
else:
silhouette_scores.append(-1)
# Set a negative score if there's only one
cluster
# Find the min_samples value with the maximum
silhouette score
best_min_samples=min_samples_range[np.a
rgmax(silhouette_scores)]
best_silhouette_score=max(silhouette_sc
ores)
print(f"The best value for min_samples
is {best_min_samples} with a silhouette
score of {best silhouette score}")

```

Hasilnya

```

The best value for min_samples is 9
with a silhouette
score of 0.42590274518553106

```

Gambar 7. Tampilan hasil *silhouette score*

Hasil pada tahapan sebelumnya dengan metode DBSCAN ini menghasilkan *min_sample* 9 dengan nilai *Silhouette Score* 0.43.

Hasil Cluster Algoritma DBSCAN

1. Pada tahap ini, nilai epsilon (eps) dan *min samples* yang telah diperoleh dari pengujian sebelumnya diinput ke dalam model. Proses klusterisasi dilakukan terhadap data dalam *DataFrame*, di mana hasil pemberian label kemudian disimpan ke dalam kolom baru bernama *Cluster_DBSCAN* (atau sesuai nama kolom yg dibuat). Langkah terakhir adalah menampilkan hasil pengelompokan yang

mencakup identitas nama negara beserta label *cluster* masing-masing. *Script* dan hasilnya ditampilkan pada gambar 8.

```

#Inisialisasi dan fitting DBSCAN
DBSCAN = DBSCAN(eps=0.17,min_samples=9)
DBSCAN_clusters=DBSCAN.fit_predict(x_sc
aled)
#Add clustering results to theDataFrame
data['DBSCAN_Cluster']= DBSCAN_clusters
# Display results
print(data[['NEGARA', 'DBSCAN_Cluster']]
)

```

Hasilnya:

	NEGARA	DBSCAN_Cluster
0	REP.RAKYAT CINA	5
1	AMERIKA SERIKAT	0
2	INDIA	-1
3	JEPANG	0
4	PILIPINA	0
..
170	MACAU	1
171	REP.AFRIKA TENGAH	2
172	SAMOA AMERIKA	-1
173	MONTENEGRO	6
174	FD STS MICRONESIA	2

[175 rows x 2 columns]

Gambar 8. Tampilan hasil cluster yang dihasilkan

2. Menampilkan hasil *clustering* dari tahapan sebelumnya

```

from collections import Counter
# Hitung jumlah data dalam setiap
cluster
cluster_counts=Counter(DBSCAN_clusters)
# Cetak hasilnya
for cluster, count in cluster_counts.items(
):
print(f"Cluster {cluster}: {count} data
points")

```

Hasilnya:

```

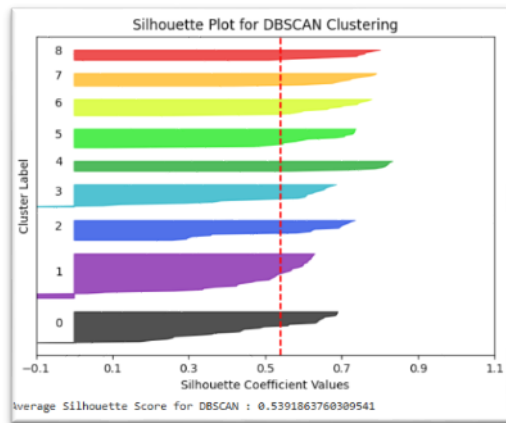
Cluster 5: 16 data points
Cluster 0: 26 data points
Cluster -1: 17 data points
Cluster 3: 19 data points
Cluster 1: 37 data points
Cluster 2: 17 data points
Cluster 8: 9 data points
Cluster 4: 9 data points
Cluster 7: 11 data points
Cluster 6: 14 data points

```

Gambar 9. Tampilan hasil *Cluster*

3. Menghitung nilai *Silhouette Coefficient* untuk memvalidasi pengelompokan. Berdasarkan visualisasi grafik hasil *clustering* algoritma DBSCAN, sumbu-y menghasilkan berbagai *cluster* yang terbentuk, termasuk beberapa kelompok berskala kecil yang diidentifikasi

melalui perbedaan warna. Titik data dengan label '-1' (ditampilkan dalam warna abu-abu) dikategorikan sebagai noise points atau pencilan yang tidak memenuhi kriteria kepadatan *cluster* mana pun. Garis merah vertikal pada grafik menunjukkan nilai rata-rata *Silhouette Score* sebesar 0,53. Berikut gambar 10 hasil perhitungan *Silhouette Score*.



Gambar 10. Hasil *Silhouette Score* Algoritma DBSCAN

Pembahasan

Dari tahap preprocessing data yang memastikan bahwa data yang digunakan sesuai dengan format yang diinginkan, dengan itu proses pembersihan dan normalisasi data dilakukan pada data ekspor non-migas dengan jumlah 175 negara. Pada algoritma DBSCAN menghasilkan 9 *cluster* dan 1 *cluster* dengan label "-1" mewakili *noise points* yang mencakup 17 negara. Pengelompokan *cluster* dianalisis berdasarkan tiga kategori tren pertumbuhan, yaitu Tren Positif Tinggi (> 20%), Tren Positif Sedang (5%–20%), dan Tren Rendah (\leq 5% atau menurun). Kelompok dengan anggota terbesar adalah *Cluster* 1 (37 negara) dan *Cluster* 0 (26 negara). Pada *Cluster* 1, dominasi Tren Positif Tinggi diwakili oleh negara seperti Taiwan dan Polandia, Tren Positif Sedang didominasi oleh Belanda dan Australia. *Cluster* 4 dan *Cluster* 8 merupakan kelompok terkecil yang masing-masing 9 negara dengan variasi tren yang ketat (misalnya, Swiss dan Angola pada Tren Positif Tinggi, serta Rusia dan Yaman pada Tren Positif Rendah). *Cluster* lainnya menunjukkan sebaran yang variatif; *Cluster* 3 (19 negara) didominasi oleh Tren Positif Sedang seperti Korea Selatan dan Inggris, *Cluster* 2 (17 negara), *Cluster* 5 (16 negara), *Cluster* 6 (14 negara), dan *Cluster* 7 (11 negara).

Kesimpulan Dan Saran

Kesimpulan

Pada hasil dan pembahasan pada bab sebelumnya, dapat disimpulkan bahwa Melalui visualisasi metode *Elbow* pada algoritma DBSCAN, parameter optimal yang diperoleh melalui pengujian sistematis adalah nilai *Epsilon* (ϵ) sebesar 0,17 dengan *MinPts* sebesar 9.

Hasil Pengelompokan Data, Algoritma DBSCAN: Menghasilkan 9 kluster utama serta satu kelompok label "-1" (*noise points*). Sebaran data terdiri dari Kluster 0 (26 negara), *Cluster* 1 (37 negara), *Cluster* 2 (17 negara), *Cluster* 3 (19 negara), *Cluster* 4 (9 negara), *Cluster* 5 (16 negara), *Cluster* 6 (14 negara), *Cluster* 7 (11 negara), *Cluster* 8 (9 negara), label "-1" ada 17 negara yang teridentifikasi sebagai *noise points*. Dan pengujian *Silhouette Score*, algoritma DBSCAN menghasilkan skor 0,53, yang mengindikasikan struktur *Cluster* yang memadai, DBSCAN menunjukkan keunggulan dalam mendeteksi data noise serta mengidentifikasi kelompok dengan pola distribusi yang tidak beraturan.

Saran

Untuk penelitian selanjutnya, diharapkan menggunakan metode baru dengan algoritma OPTICS. Dikarenakan algoritma ini dapat menyempurnakan metode DBSCAN yang dirancang khusus untuk mengatasi struktur kluster dengan tingkat kepadatan yang beragam.

Daftar Pustaka

- [1] N. Arminarahmah, A. G. Daengs, J. Tata Hardinata, And I. Kalimantan Muhammad Arsyad Al Banjari, "Klusterisasi Impor Beras Di Indonesia Menurut Negara Asal Utama Menggunakan Algoritma K-Medoids," *Jurnal Riset Sistem Informasi Dan Teknik Informatika (Jurasic)*, Vol. 8, No. 2, Pp. 793–801, 2023, [Online]. Available: <https://Tunasbangsa.Ac.Id/Ejurnal/Index.Php/Jurasic>
- [2] I. Paujiyah, "Analisis Bibliometrik Keterkaitan Ekspor Terhadap Pertumbuhan Ekonomi Indonesia," *Sanskara Manajemen Dan Bisnis*, Vol. 2, No. 01, Pp. 16–23, Nov. 2023.
- [3] A. Rezandy And A. Yasin, "Pengaruh Nilai Tukar, Inflasi, Dan Pendapatan Nasional Terhadap Ekspor Nonmigas Indonesia," *Independent : Journal Of Economics*, Vol. 1, Pp. 95–110, 2021, [Online]. Available: <https://Ejournal.Unesa.Ac.Id/Index.Php/Independent>
- [4] E. A. Fadlilah, "Identifikasi Anomali Data Akademik Menggunakan Dbscan Outlier Detection," *Prosiding Sains Nasional Dan Teknologi*, Vol. 12, No. 1, P. 336, Nov. 2022, Doi: 10.36499/Psnst.V12i1.7012.
- [5] E. Rahmah, E. Haerani, A. Nazir, S. Ramadhani, And T. Informatika Uin Sultan Syarif Kasim Riau Jl Soebrantas Km, "Penerapan Algoritma K-Medoids Clustering Untuk Menentukan Strategi Promosi Pada Data Mahasiswa (Studi Kasus : Stikes Perintis Padang)," *Jurnal Nasional Komputasi Dan Teknologi Informasi*, Vol. 5, No. 3, 2022.
- [6] B. Riyanto, "Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokan Penyebaran Diare Di Kota Medan (Studi Kasus: Kantor Dinas Kesehatan Kota Medan)," *Komik (Konferensi Nasional Teknologi Informasi Dan Komputer)*, Vol. 3, No. 1, Dec. 2019, Doi: 10.30865/Komik.V3i1.1659.

- [7] A. Gede, S. Pradnyana, M. Kom, K. Kom, S. Agustini, And M. S. Si, “Konsep Dasar Data Mining.”
- [8] F. Sains Dan Teknologi -Universitas Kanjuruhan Malang Bagus Wira, B. Wira, A. Endy Budianto, And A. Sartika Wiguna, “Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang,” 2019.
- [9] R. R. Muhima *Et Al.*, *Kupas_Tuntas_Algoritma_Clustering_Konsep*. 2022.
- [10] A. Atira And B. Nurina Sari, “Penerapan Silhouette Coefficient, Elbow Method Dan Gap Statistics Untuk Penentuan Cluster Optimum Dalam Pengelompokan Provinsi Di Indonesia Berdasarkan Indeks Kebahagiaan,” *Jurnal Ilmiah Wahana Pendidikan*, Vol. 9, No. 17, Pp. 76–86, 2023, Doi: 10.5281/Zenodo.8282638.
- [11] T. D. Harjanto, A. Vatesia, And R. Faurina, “Analisis Penetapan Skala Prioritas Penanganan Balita Stunting Menggunakan Metode Dbscan Clustering,” 2021. [Online]. Available: [Http://Ejournal.Unib.Ac.Id/Index.Php/Rekursif/30](http://ejournal.unib.ac.id/index.php/rekursif/30)
- [12] U. Hayati, “Clustering Produk Ekspor Indonesia Berdasarkan Tingkat Permintaan Menggunakan Metode K-Means Tahun 2020-2022,” 2023.
- [13] B. Nurina Sari, A. Primajaya, And J. H. Ronggowaluyo Teluk Jame Karawang, “Penerapan Clustering Dbscan Untuk Pertanian Padi Di Kabupaten Karawang,” 2019. [Online]. Available: [Www.Mapcoordinates.Net/En](http://www.mapcoordinates.net/en).
- [14] R. Adha, N. Nurhaliza, And U. Soleha, “Perbandingan Algoritma Dbscan Dan K-Means Clustering Untuk Pengelompokan Kasus Covid-19 Di Dunia,” *Jurnal Sains, Teknologi Dan Industri*, Vol. 18, No. 2, Pp. 206–211, 2021, [Online]. Available: [Https://Covid19.Who.Int](https://covid19.who.int).
- [15] M. M. Putri, C. Dewi, E. Permata Siam, G. Asri Wijayanti, N. Aulia, And R. Nooraeni, “Comparison Of Dbscan And K-Means Clustering For Grouping The Village Status In Central Java 2020 Komparasi Dbscan Dan K-Means Clustering Pada Pengelompokan Status Desa Di Jawa Tengah Tahun 2020,” Vol. 17, No. 3, Pp. 394–404, 2021, Doi: 10.20956/J.V17i3.11704.
- [16] D. Armiady, “Analisis Metode Dbscan (Density-Based Spatial Clustering Of Application With Noise) Dalam Mendeteksi Data Outlier,” *Jurikom (Jurnal Riset Komputer)*, Vol. 9, No. 6, P. 2158, Dec. 2022, Doi: 10.30865/Jurikom.V9i6.5080.