

ANALISA VOLUME PENYEBARAN SAMPAH DI KARAWANG MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING

Farras Salsabila^{1*}, Taufik Ridwan², Hannie³

^{1,2,3} Universitas Singaperbangsa Karawang; Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361, Telp. (0267) 641177

Riwayat artikel:

Received: 23 Maret 2024

Accepted: 30 Maret 2024

Published: 2 April 2024

Keywords:

Sampah, Clustering, K-Means, KDD.

Correspondent Email:

2010631250012@student.unsika.ac.id

Abstrak. Pada tahun 2022, Indonesia berada pada urutan ke-4 teratas sebagai negara yang memiliki jumlah populasi tinggi. Dengan jumlah populasi yang tinggi ini, telah memunculkan sejumlah permasalahan sosial, ekonomi, dan lingkungan. Salah satu isu yang sangat terkait dengan tren ini adalah persoalan sampah. Karawang sendiri, yang merupakan salah satu kota di Jawa Barat menempati urutan ke-7 sebagai kota yang memiliki volume sampah terbanyak. Oleh karena itu, implementasi algoritma K-Means Clustering dapat membantu dalam membagi wilayah Kecamatan berdasarkan volume penyebaran sampah. Hasil penelitian ini, didapatkan hasil 2 cluster dengan cluster 0 dengan kriteria wilayah yang memiliki volume penyebaran sampah tinggi berjumlah 6 Kecamatan, sedangkan cluster 1 yang berjumlah 24 Kecamatan memiliki kriteria wilayah dengan volume penyebaran sampah rendah. Hasil evaluasi cluster menggunakan DBI didapatkan nilai sebesar 0.869, sementara itu hasil evaluasi menggunakan Silhouette didapatkan nilai sebesar 0.591. Hal ini mengindikasikan bahwa cluster yang dihasilkan memiliki kualitas yang cukup kuat.

Abstract. In the year 2022, Indonesia ranks as the 4th most populous country in the world. This high population has led to a number of social, economic and environmental problems. One issue that is closely related to this trend is waste. Karawang alone, which is one of the cities in West Java, ranks 7th as the city with the highest volume of waste. Therefore, the implementation of the K-Means Clustering algorithm can help in dividing the sub-district area based on the volume of waste distribution. The results of this study, obtained the results of 2 clusters with cluster 0 with criteria for areas that have a high volume of waste distribution totaling 6 sub-districts, while cluster 1 which totals 24 sub-districts has criteria for areas with low waste distribution volume. The results of cluster evaluation using DBI obtained a value of 0.869, while the evaluation results using Silhouette obtained a value of 0.591. This indicates that the resulting cluster has a fairly strong quality.

1. PENDAHULUAN

Penanganan sampah menjadi salah satu tantangan utama dalam pengelolaan lingkungan di banyak wilayah. Indonesia, sebagai salah satu negara dengan populasi terbesar di dunia, juga dikenal sebagai salah satu negara penghasil sampah terbesar di dunia. Dalam laporan menurut *World Bank*, Indonesia menempati

peringkat ke-5 sebagai negara penghasil sampah terbanyak di dunia.

Dampak dari fenomena ini sangat dirasakan di banyak wilayah perkotaan di Indonesia. Salah satunya adalah Karawang, sebagai salah satu kota industri terbesar yang berkembang pesat di Indonesia, tidak terkecuali dalam menghadapi masalah ini. Menurut informasi

yang diperoleh dari situs *website* resmi Kementerian Perindustrian Indonesia pada tahun 2023, Karawang telah mengembangkan total 17 kawasan industri yang mencakup luas lahan seluas 7.657,34 hektar. Pertumbuhan ini tidak hanya berdampak pada perkembangan ekonomi dan industri, tetapi juga meningkatkan masalah lingkungan, termasuk peningkatan volume sampah dan penyebarannya di sekitar kawasan industri tersebut.

Seiring dengan pertumbuhan ekonomi dan industri yang pesat, terdapat peningkatan volume sampah yang tidak terelakkan. Pada tahun 2022, Karawang menjadi kota dengan urutan ke-7 tertinggi dalam jumlah timbunan sampah terbanyak di provinsi Jawa Barat, yaitu sebesar 366.436 ton sampah. Data ini diambil dari *website* resmi Kementerian Lingkungan Hidup dan Kehutanan.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengelompokkan wilayah kecamatan berdasarkan volume sampah menggunakan algoritma *K-Means Clustering*. *Clustering* merupakan salah satu metode Data Mining yang menggunakan teknik *unsupervised Learning* [1]. *Unsupervised learning* tidak memerlukan data yang telah diberi label, sehingga algoritma dapat menemukan pola yang tersembunyi atau mengelompokkan data tanpa intervensi manusia [2].

Beberapa penelitian sebelumnya yang relevan adalah penelitian mengenai pengelompokan daerah penyumbang sampah di Surabaya yang dilakukan oleh Sylvia dan Lukamndono. Data yang diambil adalah Data sampah masuk, sampah terkelola, sampah daur ulang, dan sampah *up-cycle*. Hasil yang diperoleh adalah jumlah kluster adalah 2, dengan nilai DBI sebesar 0.268, mengelompokkan 11 wilayah kedalam *cluster* tinggi [3].

Penelitian mengenai pengelompokan kecamatan di Yogyakarta berdasarkan jenis sampah dilakukan oleh Meimun, dkk. Data yang diambil diantaranya Jumlah sampah plastik, jumlah sampah kertas, jumlah sampah kaca, dan jumlah sampah logam. Dengan menggunakan 5 metode gabungan, metode paling baik adalah *Average Linkage* dengan nilai tertinggi. Kluster yang dihasilkan berjumlah 3 kluster, dimana kluster 1 terdiri dari 6

kecamatan, kluster 2 terdiri dari 5 kecamatan, dan kluster 3 terdiri dari 1 kecamatan [4].

Penelitian yang dilakukan oleh Tyas & Purnamasari mengenai pengelompokan kasus DBD berdasarkan Kabupaten, menghasilkan jumlah kluster optimal adalah 3. Data yang diolah adalah Nama Kabupaten/Kota, jumlah kasus, tahun. Nilai DBI yang dihasilkan sebesar 0,359. Jumlah wilayah dalam *cluster* 0 (rendah) adalah 100, dalam *cluster* 1 (tinggi) adalah 20, dan dalam *cluster* 2 (sedang) adalah 96 [5].

Dalam penelitian mengenai pengelolaan sampah ini, penerapan algoritma *Clustering K-Means* ini dapat mendukung analisis pola distribusi sampah di Kabupaten Karawang. Adapun metodologi penelitian yang digunakan adalah KDD (*Knowledge Discovery on Databases*), yang berguna sebagai metodologi dalam rancangan penelitian. Dengan metode Elbow untuk menentukan jumlah optimal *cluster* dan metode evaluasi *cluster* yang digunakan adalah DBI (*Davies Boouldin-Index*) dan diperkuat dengan metode *Silhouette Coefficient*.

2. TINJAUAN PUSTAKA

2.1 Clustering

Data mining adalah suatu proses yang dimaksudkan untuk menghasilkan informasi yang berguna dari sekumpulan data besar. Proses ini melibatkan berbagai teknik dan algoritma untuk mengeksplorasi, menganalisis, dan mengekstrak pola atau pengetahuan yang tersembunyi dalam data, dengan tujuan untuk membuat keputusan yang lebih baik dan memperoleh wawasan yang berharga dari data yang ada [6]. Salah satu metode dalam data mining adalah *Clustering*. Dalam prosesnya, *clustering* memiliki fungsi untuk mengelompokkan dataset menjadi sejumlah kelompok di mana entitas atau objek yang memiliki kesamaan atau ciri yang serupa akan dikelompokkan bersama [7].

2.2 K-Means

Algoritma *K-means* menggunakan teknik pengelompokan data non-hirarkisasi untuk memisahkan data ke dalam satu atau lebih kelompok atau kluster [8]. Tujuan dari *K-means* sendiri adalah dipergunakan untuk mengelompokkan data yang menunjukkan kesamaan atau pola yang serupa ke dalam

kelompok-kelompok tertentu [9]. Dalam langkah awalnya, metode ini menginisiasi titik-titik awal untuk setiap kluster yang disebut sebagai Centroid, yang dipilih secara acak dari data yang ada. Kemudian, prosesnya berlanjut dengan mengulangi perhitungan untuk menemukan posisi optimal dari setiap *Centroid* [2].

2.3 Cluster Validation Index

Cluster Validation Index merupakan metrik yang digunakan untuk mengevaluasi seberapa baik sebuah partisi data telah dikelompokkan menjadi kluster atau kelompok. Tujuan dari indeks validitas adalah untuk mendapatkan jumlah kluster optimal, dan metode yang digunakan oleh indeks validitas adalah dengan mengevaluasi hasil dari algoritma pengelompokan [10]. Metode yang termasuk dalam *cluster validation index*, diantaranya ada DBI (*Davies Bouldin-Index*) dan *Silhouette Coefficient*.

2.3.1 DBI

Indeks Davies-Bouldin (DBI) adalah alat evaluasi yang dapat digunakan sebagai salah satu faktor untuk memilih model atau konfigurasi terbaik dalam algoritma pengelompokan [7]. Nilai DBI yang semakin rendah, atau mendekati nilai 0, maka kualitas kluster yang dihasilkan semakin baik pula [11].

2.3.2 Silhouette Coefficient

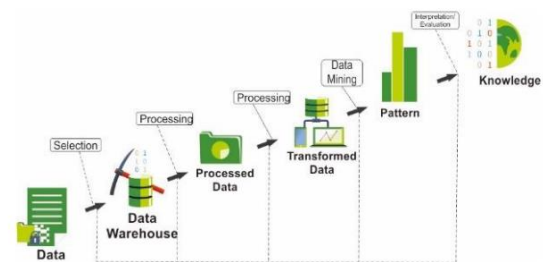
Silhouette coefficient berperan sebagai indikator untuk mengevaluasi kualitas kluster yang terbentuk, sambil juga menunjukkan seberapa baik setiap objek cocok dengan kluster tempat mereka ditempatkan [12]. Rentang nilai silhouette berkisar dari 1 hingga -1. Semakin kecil jarak antara objek dalam satu kelompok dan semakin besar jarak antara objek dengan objek di kelompok lain, maka semakin baik kualitas kluster tersebut [13].

Table 1. Rentang Nilai Silhouette dan Interpretasi [12]

Rentang Nilai	Interpretasi
0.71 – 1.00	Kuat
0.51 – 0.70	Beralasan/Cukup Kuat
0.26 – 0.50	Lemah
≤ 0.25	Tidak ditemukannya perbedaan antar kluster

3. METODE PENELITIAN

KDD (*Knowledge Discovery in Databases*) adalah suatu metodologi untuk menemukan pengetahuan dalam *dataset*. KDD dapat diartikan sebagai langkah-langkah dalam mengekstraksi atau mengidentifikasi pola, pengetahuan, dan informasi dari dataset yang besar. Pengetahuan dan informasi yang dihasilkan melalui proses KDD bersifat inovatif, dapat dipahami dengan mudah, dan memberikan manfaat bagi para pengguna informasi[14].



Gambar 1. Tahapan Metodologi KDD [15]

Secara umum, metodologi KDD ini terdiri dari 5 proses atau tahapan didalamnya, yang dilakukan secara berurutan [16] sebagai berikut.

3.1. Data Selection

Pada tahap awal ini memfokuskan kepada pemilihan dataset yang cocok atau relevan digunakan yang sesuai dengan tujuan penelitian.

3.2. Preprocessing Data

Dalam pengertian lain, preprocessing data sama dengan cleaning data bertujuan untuk mengatasi nilai yang hilang dalam rekaman data dan menghapus gangguan atau ketidakpastian yang ada dalam data [16]. Biasanya dalam tahap ini, data akan dibersihkan dari *noise*, *outlier* maupun duplikasi atau redundansi data.

3.3. Data Transformation

Tahap transformasi data ini melibatkan modifikasi dan penggabungan data ke dalam format yang cocok untuk langkah analisis berikutnya. Tipe data diubah dan disesuaikan dengan *value* yang ada dalam *dataset*.

3.4. Data Mining

Tahap ini merupakan tahap yang paling krusial yang merujuk pada fase spesifik dalam proses KDD yang bertumpu pada pemanfaatan

teknik dan algoritma data mining untuk mengurai pola, relasi, dan informasi tersembunyi dalam kumpulan data.

3.5. Evaluation

Pada tahap ini, juga diperiksa kecocokan pola atau data yang telah dianalisis dengan fakta atau asumsi yang telah ada sebelumnya [17]. Metode evaluasi dalam penelitian ini memakai 2 metode yakni DBI, dan Silhouette Coefficient.

3.6. Knowledge

Tahap ini melibatkan analisis mendalam terhadap pola, hubungan, atau tren yang teridentifikasi dari data untuk membuat kesimpulan yang relevan dan mengambil keputusan yang informasinya didukung berdasarkan hasil analisa yang didapatkan pada beberapa tahap sebelumnya.

4. HASIL DAN PEMBAHASAN

Tahapan penelitian ini mengacu pada metodologi yang digunakan, yaitu metodologi KDD (*Knowledge Discovery in Databases*).

4.1 Data Selection

Dataset yang akan diolah merupakan data pengelolaan sampah Kabupaten Karawang pada tahun 2021 dan 2022. Data ini merupakan data primer yang diambil langsung dari Dinas Lingkungan Hidup dan Kehutanan (DLHK) Karawang. Dalam tahap ini, dataset akan diseleksi berdasarkan atribut. Atribut yang sesuai dan relevan dengan tujuan penelitian akan diolah lebih lanjut. Adapun hasil akhir data selection dapat dilihat pada **Gambar 2**. Atribut yang dipakai disini menggunakan 5 atribut, yaitu nama kecamatan, volume sampah, volume sampah tertangani, pengurangan sampah, dan jumlah armada.

NO	KECAMATAN	2021				2022			
		VOLUME SAMPAH	VOLUME SAMPAH TERTANGANI	PENGURANGAN SAMPAH	JUMLAH ARMADA	VOLUME SAMPAH	VOLUME SAMPAH TERTANGANI	PENGURANGAN SAMPAH	JUMLAH ARMADA
1	PANGKALAN	7154	0	0	0	5974	0	0	0
2	TEGALWARU	7016	0	0	0	5907	0	0	0
3	CIANJEL	8256	0	0	0	6874	0	0	0
4	TELUKANJEBE TIMUR	25077	17554	7523	10	20707	38897	1807	17
5	TELUKANJEBE BARAT	10254	7041	3213	6	8546	8450	72	1
6	KLARI	33839	22203	11636	1	38541	10390	176	2
7	CIRAMPEK	21407	14728	6679	4	17737	1129	176	3
8	PURWASARI	14506	9150	5356	1	12128	3538	72	1
9	TERTAMPAJA	9227	5803	3424	1	7946	1095	0	1
10	JATISARI	14524	944	5084	2	12000	2581	25	1
11	BANTUSARI	10301	649	3811	1	8580	0	0	0
12	KOTABARU	24413	16747	7666	0	20482	2975	54	2
13	CEAMAYA WETAN	14483	9124	5359	1	12055	1117	23	1
14	CEAMAYA KULON	12031	758	4451	0	9978	1095	0	1
15	LEMMA ABANG	12011	7587	4444	1	9915	1095	0	1
16	TELAGA SARI	12554	816	4394	2	10375	2223	32	1
17	MAJALAYA	12274	815	4124	0	10732	1621	518	1
18	KARAWANG TIMUR	2736	1866	87	5	22780	12631	940	7
19	KARAWANG BARAT	29804	20863	8941	20	24635	17819	655	11
20	RAWAMERTA	9929	0	0	2	8197	0	0	0
21	TEMPURAN	11988	0	0	1	9997	0	0	0
22	KUTAWALITYA	11024	7055	3969	1	9180	0	0	0
23	PENGASDENKLOK	20409	13877	6531	5	17095	11125	173	7
24	JAYAKERTA	12146	0	0	1	10169	0	0	0
25	PEDES	14572	0	0	1	12091	0	0	0
26	CEBAR	8102	0	0	0	6699	0	0	0
27	CIBUAYA	968	0	0	0	8083	0	0	0
28	TIRTAYAYA	13025	0	0	0	10895	0	0	0
29	BATUJAYA	14549	0	0	0	12279	0	0	0
30	PAKSIJAYA	7453	0	0	0	2641	0	0	0

Gambar 2. Hasil Data Selection

4.2 Preprocessing Data

Pengecekan *missing value* dan duplikasi data dilakukan pada tahap ini. Sesuai dengan tujuannya, nantinya data yang tidak normal atau tidak sesuai akan ditindak lanjut, ini bisa data tersebut dihilangkan atau diperbaiki.

4.3 Data Transformation

Pada tahap ini, nilai atau isi dalam dataset akan diubah kedalam format yang sesuai, sehingga data dapat diolah dengan baik.

```
data_select['KECAMATAN'] =
data_select['KECAMATAN'].astype('string')
data_select['VOLUME SAMPAH 2021'] =
data_select['VOLUME SAMPAH
2021'].astype('int')
data_select['VOLUME SAMPAH 2022'] =
data_select['VOLUME SAMPAH
2022'].astype('int')
data_select['VOLUME SAMPAH TERTANGANI
2021'] = data_select['VOLUME SAMPAH
TERTANGANI 2021'].astype('int')
data_select['VOLUME SAMPAH TERTANGANI
2022'] = data_select['VOLUME SAMPAH
TERTANGANI 2022'].astype('int')
data_select['PENGURANGAN SAMPAH 2021']
= data_select['PENGURANGAN SAMPAH
2021'].astype('int')
data_select['PENGURANGAN SAMPAH 2022']
= data_select['PENGURANGAN SAMPAH
2022'].astype('int')
data_select['JUMLAH ARMADA 2021'] =
data_select['JUMLAH ARMADA
2021'].astype('int')
```

```
data_select['JUMLAH ARMADA 2022'] =
data_select['JUMLAH ARMADA
2022'].astype('int')
```

Dalam penelitian ini, data yang bersifat numerik akan diubah ke dalam bentuk integer, sedangkan data non-numerik akan diubah kedalam tipe data string. Maka, hasil yang dapat dilihat sebagai berikut.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 9 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   KECAMATAN                            30 non-null     string
1   VOLUME SAMPAH 2021                   30 non-null     int64
2   VOLUME SAMPAH 2022                   30 non-null     int64
3   VOLUME SAMPAH TERTANGANI 2021        30 non-null     int64
4   VOLUME SAMPAH TERTANGANI 2022        30 non-null     int64
5   PENGURANGAN SAMPAH 2021             30 non-null     int64
6   PENGURANGAN SAMPAH 2022             30 non-null     int64
7   JUMLAH ARMADA 2021                   30 non-null     int64
8   JUMLAH ARMADA 2022                   30 non-null     int64
dtypes: int64(8), string(1)
```

Gambar 3. Hasil Data Transformation

4.4 Data Mining

Pada tahap ini, dataset akan diolah lebih lanjut menggunakan Algoritma *K-Means Clustering*, yang bertujuan untuk mengelompokkan wilayah Kecamatan berdasarkan volume penyebaran sampah. Langkah awal dalam algoritma *K-Means Clustering* ini adalah menentukan jumlah *cluster* terlebih dahulu. Penentuan jumlah *cluster* ini bisa menggunakan macan-macam metode, salah satunya yang paling banyak digunakan adalah Metode Elbow.

Metode Elbow melibatkan plot SSE (*Sum of Square Errors*) terhadap jumlah kluster yang berbeda. Jumlah klusteer yang optimal membentuk lengkungan yang menyerupai siku pada grafik plot. Metode Elbow mendapat namanya karena kurva plot yang dihasilkannya menunjukkan lengkungan yang menyerupai siku pada bagian siku [18].

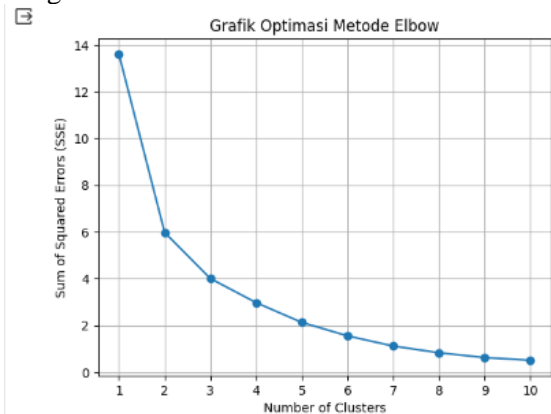
```
#Menghitung Nilai SSE
clusters = []
for i in range(1, 11):
    kmeans=KMeans(n_clusters=i, init='k-
means++', random_state=42, n_init=10)
    kmeans.fit(df_transform)
    clusters.append(kmeans.inertia_)
    print('K : ' + str(i) + ', SSE : ' +
str(kmeans.inertia_))
```

Nilai SSE dihitung untuk nilai kluster 1 – 10, dan didapatkan hasil sebagai berikut.

```
K : 1, SSE : 13.607806639746547
K : 2, SSE : 5.9747120275792955
K : 3, SSE : 4.0008591200616985
K : 4, SSE : 2.969575366654836
K : 5, SSE : 2.1224642389819515
K : 6, SSE : 1.5459902562115848
K : 7, SSE : 1.1103568046508372
K : 8, SSE : 0.8229795185835456
K : 9, SSE : 0.6170939776385092
K : 10, SSE : 0.5027759468030876
```

Gambar 4. Perhitungan SSE

Setelah diketahui nilai SSE, selanjutnya dibuat grafik Elbow untuk memvisualisasikan hubungan antara jumlah kluster yang berbeda dengan nilai SSE.



Gambar 5. Grafik Elbow

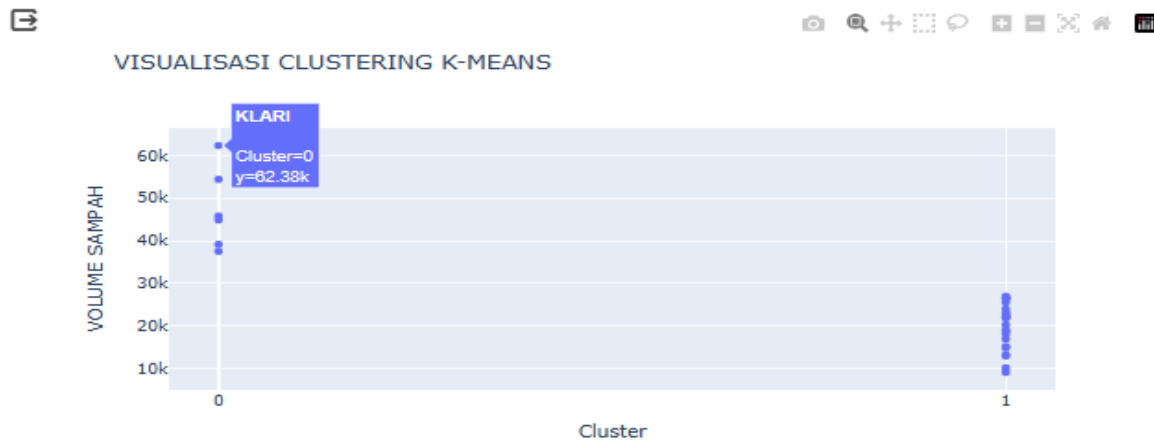
Pada Gambar 5., titik yang membentuk siku ada pada *cluster* 2. Hal ini mengindikasikan bahwa jumlah optimal *cluster* adalah 2 *cluster*. Selanjutnya, masuk kedalam proses *K-Means Clustering*. Langkah awalnya adalah menghitung nilai centroid untuk masing-masing kluster berdasarkan data yang tersedia.

```
[[0.75618833 0.72941441 0.78709505 0.35379142 0.71565514 0.27606242
0.33333333 0.41176471]
[0.28868278 0.27389801 0.09572057 0.0379939 0.15725344 0.03577357
0.05416667 0.03921569]]
array([1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1,
0, 1, 1, 1, 1, 1, 1], dtype=int32)
```

Gambar 6. Hasil Centroid

Dalam Gambar 6., 2 baris awal adalah nilai *centroid*, sementara baris terakhir menampilkan label kluster untuk setiap sampel data. Angka dalam label kluster (0 dan 1) mengindikasikan kluster mana setiap sampel data termasuk di dalamnya.

Untuk dapat memahami lebih mudah, maka dari hasil proses *K-Means Clustering*, dibuatlah visualisasi grafiknya.



Gambar 7. Grafik Visualisasi K-Means Clustering

Dalam visualisasi tersebut, *cluster 0* memiliki posisi yang paling tinggi dibandingkan *cluster 1*. Oleh karena itu, ini membuktikan bahwa wilayah dengan volume penyebaran sampah tinggi adalah wilayah kecamatan yang termasuk anggota dari *cluster 0*. Sementara itu, wilayah dengan penyebaran volume sampah rendah merupakan anggota dari *cluster 1*.

4.5 Evaluation

Pada tahap evaluasi ini, kluster yang sudah dihasilkan akan dicek bagaimana kualitasnya. Metode DBI ini dapat mengetahui seberapa baik atau kuatnya suatu kluster dihasilkan.

```
#DBI
from sklearn.cluster import KMeans
from sklearn.metrics import davies_bouldin_score
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
data_standardized = scaler.fit_transform(df_transform)

# Menetapkan nilai n_init secara eksplisit
kmeans = KMeans(n_clusters=2, random_state=42, n_init=10)
predicted_labels = kmeans.fit_predict(data_standardized)

# Kalkulasi Davies-Bouldin Index
dbi = davies_bouldin_score(data_standardized, predicted_labels)

print(f"Davies-Bouldin Index: {dbi}")
```

Davies-Bouldin Index: 0.8697313927631934

Gambar 8. Perhitungan Nilai DBI

Dari perhitungan DBI diatas, nilai yang didapat sebesar 0.869. Dalam aturan DBI, semakin nilai DBI mendekati 0 atau semakin rendah nilainya, maka semakin baik pula *cluster* yang dihasilkan [11]. Untuk memperkuat hasil evaluasi ini, maka dalam penelitian ini juga memakai metode *Silhouette Coefficient*.

```
#SILHOUETTE COEFFICIENT
from sklearn.metrics import silhouette_score

km = KMeans(n_clusters=2, init='k-means++',
            n_init=10, max_iter=100,
            random_state=None)
km.fit_predict(df_transform)

score = silhouette_score(df_transform,
                        km.labels_, metric='euclidean')
print('Silhouette Score: %.3f' % score)
```

Silhouette Score: 0.591

Gambar 9. Perhitungan Nilai Silhouette

Nilai *Silhouette* yang didapat adalah 0.591., nilai tersebut masuk kedalam rentang nilai *Silhouette Coefficient*, yakni berkisar antara -1 sampai dengan 1.

4.6 Knowledge

Tahap terakhir ini adalah tahap dimana informasi diperoleh dari data yang sudah diekstrak sebelumnya. Telah diketahui sebelumnya, bahwa jumlah kluster yang optimal adalah 2 *cluster*. Dimana, *cluster 0* merupakan wilayah dengan jumlah penyebaran sampah tinggi, sedangkan *cluster 1* merupakan wilayah dengan jumlah penyebaran sampah tinggi. Untuk memudahkan dalam memahami dan melihat hasil pengelompokkan, maka dibuatlah suatu tabel sebagai berikut dengan memasukkan kolom Kecamatan, *cluster*, dan kriteria.

Table 2. Hasil Pengelompokkan Cluster

No	Nama Kecamatan	Cluster	Kriteria
1	Pangkalan	1	Rendah
2	Tegalwaru	1	Rendah

3	Ciampel	1	Rendah
4	Telukjambe Timur	0	Tinggi
5	Telukjambe Barat	1	Rendah
6	Klari	0	Tinggi
7	Cikampek	0	Tinggi
8	Purwasari	1	Rendah
9	Tirtamulya	1	Rendah
10	Jatisari	1	Rendah
11	Banyusari	1	Rendah
12	Kotabaru	0	Tinggi
13	Cilamaya Wetan	1	Rendah
14	Cilamaya Kulon	1	Rendah
15	Lemah Abang	1	Rendah
16	Telagasari	1	Rendah
17	Majalaya	1	Rendah
18	Karawang Timur	1	Rendah
19	Karawang Barat	0	Tinggi
20	Rawamerta	1	Rendah
21	Tempuran	1	Rendah
22	Kutawaluya	1	Rendah
23	Rengasdengklok	0	Tinggi
24	Jayakarta	1	Rendah
25	Pedes	1	Rendah
26	Cilebar	1	Rendah
27	Cibuaya	1	Rendah
28	Tirtajaya	1	Rendah
29	Batujaya	1	Rendah
30	Pakisjaya	1	Rendah

Jumlah anggota *cluster* 0 adalah 6 anggota, sementara itu jumlah anggota *cluster* 1 adalah 24 anggota. Dari hasil evaluasi *cluster*, didapatkan nilai DBI dan Silhouette, masing-masing nilai tersebut dapat dilihat pada Table 3.

Table 3. Nilai Hasil Evaluasi

DBI	0,86
Silhouette	0,591

Berdasarkan pada Table 1, nilai evaluasi *cluster* yang dihasilkan dalam penelitian ini menunjukkan bahwa kualitas *cluster* yang dihasilkan cukup kuat. Dalam artian, nilai yang terdapat dalam *dataset* menunjukkan kecenderungan mirip satu sama lain dan berbeda secara signifikan dari objek dalam *cluster* lain.

5. KESIMPULAN

- Perhitungan jumlah *cluster* optimal menggunakan metode Elbow didapatkan jumlah optimal adalah 2 *cluster*.
- Hasil evaluasi menggunakan metode DBI mendapatkan nilai sebesar 0.86 dan diperkuat dengan hasil metode *silhouette* sebesar 0.59, yang mana angka tersebut mengindikasikan bahwa *cluster* yang dihasilkan memiliki kualitas yang cukup kuat.
- Anggota *cluster* 1 yang memiliki kriteria wilayah dengan penyebaran sampah tinggi terdiri dari 6 Kecamatan, yaitu Telukjambe Timur, Klari, Cikampek, Kotabaru, Karawang Barat, dan Rengasdengklok. Sedangkan *cluster* 1 dengan kriteria wilayah yang memiliki penyebaran sampah rendah terdiri dari 24 Kecamatan yaitu Pangkalan, Tegalwaru, Ciampel, Telukjambe Barat, Purwasari, Tirtamulya, Jatisari, Banyusari, Cilamaya Wetan, Cilamaya Kulon, Lemah Abang, Telagasari, Majalaya, Karawang Timur, Rawamerta, Tempuran, Kutawaluya, Jayakarta, Pedes, Cilebar, Cibuaya, Tirtajaya, Batujaya, dan Pakisjaya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] H. Frigui, *Clustering: Algorithms and applications*. 2008. doi: 10.1109/IPTA.2008.4743793.
- [2] A. Wanto *et al.*, *Data Mining Algoritma & Implementasi*. Yayasan Kita Menulis, 2020.
- [3] S. E. Wardhani and Lukmandono, "Clustering Sampah Yang Dihasilkan Oleh Masyarakat Kota Surabaya Dengan Menggunakan Metode K-Means (Studi Kasus: Kota Surabaya)," *J. SENOPATI*, vol. 3, pp. 1–10, 2021.
- [4] Meimunah, I. Prawesti, S. Al Akhsan, and E. Widodo, "Analisis Clustering Sampah Yogyakarta.pdf." 2019.

- [5] T. M. M. Tyas and A. I. Purnamasari, "Penerapan Algoritma K-means dalam Mengelompokkan Demam Berdarah Dengue Berdasarkan Kabupaten," *Blend Sains J. Tek.*, vol. 1, no. 4, pp. 277–283, 2023, doi: 10.56211/blendsains.v1i4.231.
- [6] R. R. Rerung, "Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk," *J. Teknol. Rekayasa*, vol. 3, no. 1, p. 89, 2018, doi: 10.31544/jtera.v3.i1.2018.89-98.
- [7] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *J. Sains dan Manaj.*, vol. 9, no. 1, p. 96, 2021, [Online]. Available: www.bps.go.id
- [8] F. Khoirunnisa and Y. Rahmawati, "Komparasi 2 Metode Cluster Dalam Pengelompokan Intensitas Bencana Alam Di Indonesia," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 1, pp. 68–79, 2024, doi: 10.23960/jitet.v12i1.3619.
- [9] H. Kurniawan, S. Defit, and Sumijan, "Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," *J. Appl. Comput. Technol.*, vol. 1, no. 2, pp. 80–89, 2020.
- [10] A. Aditya, I. Jovian, and B. N. Sari, "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019," *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 51, 2020, doi: 10.30865/mib.v4i1.1784.
- [11] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [12] T. A. Novia, D. Swanjaya, and R. Wulaningrum, "Analisis Clustering Aplikasi Laporan Kasus Kekerasan," *J. Borneo Inform. Tek. Komput.*, vol. 2, no. 1, pp. 17–30, 2022.
- [13] K. P. Simanjuntak and U. Khaira, "Pengelompokan Titik Api di Provinsi Jambi dengan Algoritma Agglomerative Hierarchical Clustering," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 7–16, 2021, doi: 10.57152/malcom.v1i1.6.
- [14] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 20–28, 2023, doi: 10.58602/jaiti.v1i1.22.
- [15] S. Oktarian, S. Defit, and Sumijan, "Clustering Students' Interest Determination in School Selection Using the K-Means Clustering Algorithm Method," *J. Inf. dan Teknol.*, vol. 2, pp. 68–75, 2020, doi: 10.37034/jidt.v2i3.65.
- [16] M. A. Muslim *et al.*, *DATA MINING ALGORITMA C4.5*. 2019.
- [17] S. Fathuroh, "Metode K-Means Clustering Dalam Optimalisasi Kinerja Dosen Pendamping Akademik Pada Program Kampus Merdeka," *J. Sistim Inf. dan Teknol.*, vol. 5, pp. 5–9, 2022, doi: 10.37034/jsisfotek.v5i2.172.
- [18] A. R. Lashiyanti, I. Rasyid Munthe, F. A. Nasution, and E. P. Korespondensi, "Optimisasi Klasterisasi Nilai Ujian Nasional dengan Pendekatan Algoritma K-Means, Elbow, dan Silhouette," *J. Ilmu Komput. dan Sist. Inf. (JIKOMSI)*, vol. 6, no. 1, pp. 14–20, 2023.