PENERAPAN METODE *MEAN SHIFT CLUSTERING* UNTUK MENGELOMPOKKAN WILAYAH BERDASARKAN PENGELOLAAN SAMPAH

Awal Lidya Musaffak, Kartika Maulida Hindrayani, Mohammad Idhom
Sains Data, UPN "Veteran" Jawa Timur
Jl. Raya Rungkut Madya, Gunung Anyar, Surabaya
21083010088@student.upnjatim.ac.id

ABSTRAK

Pengelolaan sampah di Indonesia menjadi tantangan besar dengan meningkatnya timbulan sampah setiap tahun. Data SIPSN 2023 mencatat timbulan sampah harian sebesar 106.145,71 ton dan tahunan mencapai 38.743.185,18 ton. Setiap wilayah memiliki pola pengelolaan sampah yang berbeda, sehingga diperlukan segmentasi untuk memahami variasinya. Penelitian ini menerapkan algoritma *Mean Shift Clustering* untuk mengelompokkan wilayah berdasarkan data pengurangan dan penanganan sampah di setiap kabupaten dan kota. Dengan *bandwidth* 1.5, hasil analisis menunjukkan terbentuknya dua klaster dengan nilai *Silhouette Score* sebesar 0.649. Terdapat dua klaster yang dihasilkan dengan klaster 1 merupakan klaster dengan sampah yang terkelola rendah sedangkan klaster 2 adalah klaster dengan sampah terkelola tinggi. Hasil penelitian ini diharapkan dapat membantu dalam perumusan kebijakan yang lebih tepat sasaran untuk meningkatkan pengelolaan sampah secara efisien dan berkelanjutan di berbagai daerah.

Kata kunci : *Pengelolaan Sampah, Clustering, Mean Shift , Silhouette Score*

1. PENDAHULUAN

Sampah adalah sisa material dari aktivitas manusia yang sudah tidak digunakan dan dibuang, baik dalam bentuk padat, cair, maupun gas[1]. Sampah telah menjadi bagian tak terpisahkan dari kehidupan manusia. Seiring dengan pertumbuhan populasi, volume sampah terus mengalami peningkatan setiap harinya[2]. Pengelolaan sampah yang tidak efektif dapat menyebabkan berbagai permasalahan lingkungan, seperti pencemaran tanah, air, dan udara, serta dampak negatif terhadap kesehatan masyarakat. Kementerian Lingkungan Hidup dan Kehutanan telah meluncurkan Sistem Informasi Pengelolaan Sampah Nasional (SIPSN). Situs ini menyediakan berbagai informasi terkait pengelolaan sampah di Indonesia, termasuk data pengelolaan, regulasi, fasilitas, serta pemetaan sistem pengelolaannya.[3]

Berdasarkan data dari SIPSN tahun 2023, Indonesia menghasilkan timbulan sampah harian sebesar 106.145,71 ton dan timbulan sampah tahunan mencapai 38.743.185,18 ton[4]. Angka ini menjadikan tahun 2023 sebagai tahun dengan jumlah timbulan sampah tertinggi dibandingkan tahun-tahun sebelumnya. Pada tahun 2021, timbulan sampah harian tercatat sebesar 78.332,39 ton dengan total tahunan 28.591.323,10 ton, sedangkan pada tahun 2022, jumlah timbulan sampah harian mencapai 105.845,77 ton dengan total tahunan 38.633.706,53 ton. Peningkatan ini menunjukkan bahwa pengelolaan sampah yang lebih efektif dan terstruktur menjadi kebutuhan yang semakin mendesak.

Setiap daerah memiliki karakteristik berbeda dalam pengurangan dan penanganan sampah, yang dipengaruhi oleh faktor seperti kebijakan, infrastruktur, dan tingkat partisipasi masyarakat. Perbedaan ini berkontribusi terhadap jumlah timbulan sampah di masing-masing wilayah, yang mencerminkan volume sampah dari berbagai sumber dalam kurun waktu tertentu serta menjadi indikator efektivitas pengelolaannya. Jumlah timbulan sampah yang tinggi dapat menunjukkan sistem pengelolaan yang kurang baik, yang pada akhirnya memicu pencemaran lingkungan serta berdampak negatif pada kesehatan masyarakat. Penumpukan sampah yang tidak terkelola dengan baik juga dapat menjadi tempat berkembang biaknya berbagai penyakit.

Pengelolaan sampah merupakan proses yang terstruktur, menyeluruh, dan berkelanjutan yang mencakup aktivitas pengurangan serta penanganan sampah[5]. Pengelolaan sampah terbagi menjadi dua kategori, yaitu sampah rumah tangga dan sejenisnya, serta sampah spesifik. Sampah spesifik mencakup limbah B3, sampah akibat bencana, puing bangunan, serta sampah yang belum dapat diolah dengan teknologi saat ini[6]. Sementara itu, sampah rumah tangga dan sejenisnya mencakup timbulan sampah, pengurangan, penanganan, pendauran ulang, serta sampah yang berhasil dikelola kembali. Oleh karena itu, memahami pola pengelolaan sampah di setiap wilayah menjadi langkah penting dalam merancang strategi yang lebih efektif dan berkelanjutan.

Salah satu pendekatan yang dapat digunakan untuk menganalisis antar wilayah adalah dengan menggunakan clustering[7]. Dengan menggunakan algoritma *Mean Shift Clustering*, wilayah-wilayah dapat dikelompokkan berdasarkan kemiripan dalam pola pengelolaan sampah tanpa perlu menentukan jumlah klaster terlebih dahulu. *Mean Shift* merupakan algoritma non parametrik[8]. Algoritma ini bekerja dengan mencari area kepadatan tinggi dalam data, sehingga dapat secara adaptif membentuk kelompok yang mencerminkan kondisi pengelolaan sampah di berbagai daerah. Algoritma ini dipilih karena mampu menemukan klaster dengan bentuk yang tidak

beraturan dan dapat mengelola data dengan pola yang beragam. Selain itu, Mean Shift tidak memerlukan asumsi awal mengenai jumlah klaster seperti metode K-Means, sehingga lebih fleksibel dalam menangkap pola alami dalam data pengelolaan sampah.

Beberapa penelitian sebelumnya telah membahas mengenai metode *Mean Shift* dan pengelolaan sampah. Cinderatama et al. (2022) menggunakan algoritma *Mean Shift* untuk mengelompokkan jenis dan karakteristik ancaman jaringan pada infrastruktur kesehatan pemerintah. Dari hasil percobaan, *Mean Shift* dengan *bandwidth* = 1 menghasilkan klaster terbaik dengan *Silhouette Score* sebesar 0,5305. Sementara itu, metode K-Means dengan jumlah klaster sebanyak 4 menghasilkan *Silhouette Score* sebesar 0,4531, sedangkan DBSCAN dengan parameter Eps sebesar 0,2 dan MinPts sebanyak 3 menghasilkan *Silhouette Score* sebesar 0,3424. Dengan meningkatnya serangan siber *seperti Denial of Service, Malware, dan Phishing, Intrusion Detection System* (IDS) diperlukan untuk mendeteksi ancaman. *Teknologi Data Science* diterapkan untuk meningkatkan efektivitas IDS dalam mengidentifikasi pola serangan. Hasil penelitian ini diimplementasikan dalam aplikasi berbasis *web* [9].

Salsabila et al. (2024) menerapkan algoritma K-Means *Clustering* untuk mengelompokkan wilayah kecamatan di Karawang berdasarkan volume penyebaran sampah pada tahun 2022. Hasil analisis menghasilkan dua klaster dengan *bandwidth* 285, yaitu wilayah dengan volume sampah tinggi dan rendah. Evaluasi dengan *Davies-Bouldin Index* (DBI) memperoleh nilai 0.869, sedangkan *Silhouette Score* bernilai 0.591, yang menunjukkan bahwa hasil klasterisasi cukup kuat[10].

Rizuan et al. (2023) menggunakan algoritma Mean-Shift *Clustering* untuk mengelompokkan calon penerima Bantuan Pangan Non Tunai (BPNT) di Pekanbaru berdasarkan kriteria sosial. Hasilnya menunjukkan dua klaster terbaik dengan *bandwidth* 285 dan *Silhouette Score* 0.95. Klaster 1 (680 data) didominasi penerima dengan tempat tinggal sewa, lantai batu merah, dan sumber air ledeng, sementara klaster 2 (2 data) memiliki rumah milik sendiri, lantai keramik, dan sumber air sumur bor[8].

Berdasarkan hasil penelitian terdahulu yang menunjukkan bahwa Mean Shift mampu menghasilkan klaster dengan struktur yang jelas dan pemisahan yang optimal. Melalui penelitian ini, diharapkan dapat diperoleh segmentasi wilayah berdasarkan pola pengelolaan sampah yang dapat menjadi dasar dalam perumusan kebijakan yang lebih tepat sasaran. Selain itu, hasil segmentasi ini juga dapat membantu pemerintah dalam mengidentifikasi daerah yang memerlukan peningkatan infrastruktur atau intervensi kebijakan tertentu untuk mencapai pengelolaan sampah yang lebih optimal.

2. TINJAUAN PUSTAKA

2.1. *Pengelolaan Sampah*

Timbulan sampah adalah volume atau berat sampah yang dihasilkan dalam suatu wilayah dalam satuan waktu tertentu[11]. Pengelolaan sampah merupakan proses yang terus berlangsung, dilakukan secara terstruktur dan menyeluruh, mencakup upaya pengurangan dan penanganan limbah[12]. Pengurangan dilakukan melalui kegiatan daur ulang dan pemanfaatan kembali, sementara penanganannya mencakup pengumpulan, pengangkutan, serta pembuangan akhir guna mencegah pencemaran lingkungan[5].

2.2. *Clustering*

Clustering adalah metode pengelompokan data berdasarkan kesamaan karakteristik sehingga data dalam satu kelompok lebih mirip satu sama lain dibandingkan dengan data di kelompok lain[13][14]. *Clustering* merupakan salah satu metode dari Unsupervised Learning, yang mana dalam pengaplikasiannya tidak memerlukan adanya suatu label[15].

2.3. *Mean Shift*

Algoritma *Mean Shift* adalah metode klastering non-parametrik yang tidak memerlukan penetapan jumlah klaster di awal[16]. Algoritma ini beroperasi dengan mengidentifikasi area dengan konsentrasi data yang tinggi[17], kemudian secara bertahap menggeser setiap titik data menuju rata-rata lokal dari titik-titik di sekitarnya dalam jangkauan tertentu yang disebut *bandwidth* . Selama proses iterasi, titik pusat semakin mendekati daerah dengan kepadatan lebih tinggi hingga mencapai kestabilan, membentuk klaster di sekitar titik-titik dengan distribusi data terbesar. *Mean Shift* sangat efektif dalam mengenali klaster dengan bentuk tidak teratur dan dapat digunakan dalam berbagai aplikasi tanpa bergantung pada asumsi distribusi data tertentu[8]. Berikut merupakan langkah-langkah dalam algoritma *Mean Shift* .

a. Menentukan kernel dan *bandwidth* .

$$k(x) = \left(\frac{1}{h}\right)\varphi\left(\frac{x}{h}\right) \qquad (1)$$

Dimana:
h merupakan *bandwidth*
φ merupakan fungsi gaussian yang memungkinkan dalam menentukan bobot pada tiap titik data agar sesuai dengan jarak dari pusat klaster.

b. Inisialisasi titik pusat awal.

c. Tiap *centroid* , hitung pusat massa baru menggunakan rumus *Mean Shift* .

$$m(x) = \frac{\sum(K(x-x_i)x_i)}{\sum K(x-x_i)} \qquad (2)$$

Dimana:
$m(x)$ merupakan pusat massa baru
x merupakan pusat massa saat ini
x_i merupakan titik data dalam radius kernel yang diberikan
k merupakan fungsi kernel

d. Iterasi hingga *centroid* stabil.
e. Setelah konvergensi maka tiap *centroid* mewakili mode atau pusat kelompok data.

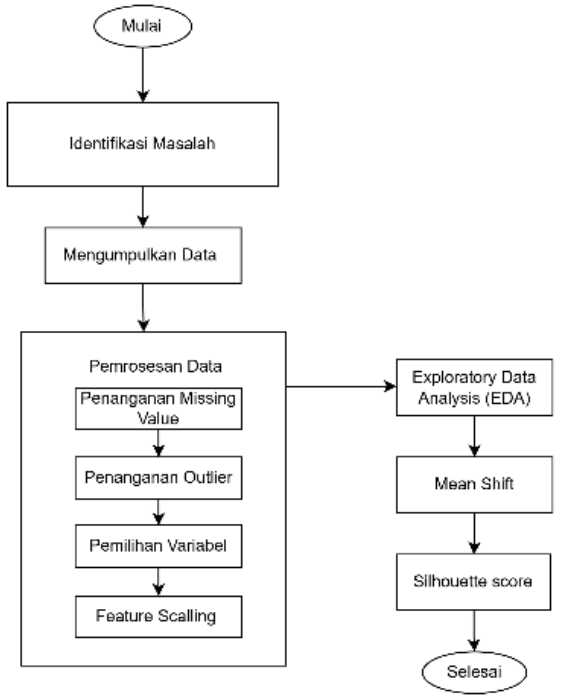
2.4. *Silhouette Score*

Silhouette Score adalah metrik yang digunakan untuk mengevaluasi kualitas hasil *Clustering* dengan mengukur seberapa baik suatu data berada dalam klusternya dibandingkan dengan kluster lain. Dalam penerapannya menggunakan perhitungan *euclidian distance*[18]. Nilai *Silhouette Score* berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa data dikelompokkan dengan baik, nilai mendekati 0 menunjukkan bahwa data berada di batas antara kluster, dan nilai negatif menunjukkan bahwa data salah pengelompokan[19]. Metrik ini sering digunakan untuk menentukan jumlah kluster yang terbentuk dalam algoritma *Clustering*[20].

$$S = \frac{b-a}{\max(a,b)} \quad (3)$$

S merupakan *silhouette score*
x merupakan pusat massa saat ini
a merupakan jarak rata-rata terhadap sampel dan semua titik lain dalam kelas yang sama
b merupakan jarak rata-rata terhadap sampel dan semua titik lain dalam klaster yang terdekat

3. METODE PENELITIAN



Gambar 1. Diagram alir metode penelitian

Metode penelitian merupakan langkah-langkah yang dilakukan dalam menerapkan algortima *Mean Shift* dalam pengelompokkan wilayah berdasarkan pengelolaan sampah. Penjelasan terkait metode penelitian tercantum dalam diagram alir pada Gambar 1.

3.1. Identifikasi Masalah

Identifikasi masalah adalah tahap awal penelitian untuk menemukan dan memahami permasalahan yang menjadi dasar kajian. Setelah itu, peneliti mendalami wawasan terkait dan menentukan metode yang tepat dengan meninjau penelitian sebelumnya. Analisis kelebihan serta kekurangan metode yang ada membantu memilih pendekatan optimal agar penelitian lebih efektif dan akurat.

3.2. Pengumpulan Data

Pengumpulan data merupakan tahap penting dalam analisis yang bertujuan untuk memperoleh informasi yang relevan sesuai dengan permasalahan penelitian. Data yang dikumpulkan harus sesuai dengan kebutuhan penelitian dan dapat diperoleh melalui berbagai metode. Dalam penelitian ini, data diambil dari *website* resmi Sistem Informasi Pengelolaan Sampah Nasional (SIPSN) dalam format *Excel*.

3.3. *Preprocessing*

Preprocessing data adalah tahap awal dalam analisis data yang bertujuan untuk membersihkan, menyusun, dan mempersiapkan data sebelum digunakan dalam pemodelan atau analisis lebih lanjut. Proses ini mencakup berbagai langkah penting, seperti menangani data yang hilang (*missing value*) dengan metode imputasimenggunakan nilai *median* untuk mengisi data yang hilang. Tahap selanjutnya adalah mengidentifikasi serta mengatasi outlier agar tidak memengaruhi hasil analisis. Dalam penelitian ini, penanganan outlier menggunakan metode Interquartile Range (IQR), di mana nilai outlier diubah agar berada dalam rentang antara kuartil 1 (Q1) dan kuartil 3 (Q3). Jika nilai outlier berada di bawah ambang batas bawah, maka nilai tersebut diubah menjadi kuartil 1 (Q1), sedangkan jika nilai outlier berada di atas ambang batas atas, maka nilai tersebut diubah menjadi kuartil 3 (Q3). Setelah itu dilakukan pemilihan variabel, dalam hali ini variabel yang akan digunakan adalah variabel sampah tahunan, pengurangan sampah, dan penanganan sampah. Lalu tahapan selanjutnya adalah *feature scaling* guna menyamakan skala data agar lebih seimbang dalam pemodelan, untuk mengatasi hal ini peneliti menggunakan metode *robust scaler*. Dengan melakukan *Preprocessing*, data menjadi lebih bersih, terstruktur, dan siap untuk dianalisis dengan lebih akurat serta menghasilkan hasil yang lebih dapat diandalkan. Dalam melakukan preprocessing ini menggunakan bahasa pemrograman *Python* serta menggunakan *library* bawaan seperti *pandas*, *matplotlib*, dan *sklearn*

3.4. *Exploratory Data Analysis* (EDA)

Tahapan *exploratory data analysis* (EDA) adalah proses analisis dan pemeriksaan dataset untuk memahami karakteristik utamanya. Langkah ini memungkinkan identifikasi pola atau informasi tersembunyi yang mungkin tidak terlihat pada tahap

pemodelan. Dalam EDA melakukan visualisasi data dengan mengunakan histogram dan *heatmap*, serta menerapkan statistika deskriptif guna memahami distribusi serta hubungan antar variabel dalam dataset.

3.5. Mean Shift Clustering

Tahapan selanjutnya merupakan tahapan pemodelan. Model yang digunakan pada penelitian ini merupakan klastering.Algoritma yang digunakan merupakan *Mean Shift*. Dalam pemodelan ini dilakukan pengukuran *bandwidth* untuk mendapatkan nilai metrik evaluasi yang terbaik.

3.6. Evaluasi

Evaluasi model merupakan tahap penting dalam analisis data untuk menilai kualitas dan kinerja suatu model. Dalam *Clustering*, evaluasi dilakukan untuk memastikan bahwa hasil pengelompokan yang dihasilkan memiliki struktur yang jelas dan representatif terhadap data yang digunakan. Salah satu metrik yang umum digunakan adalah silhouette score, yang mengukur seberapa baik suatu data berada dalam klasternya dibandingkan dengan klaster lain.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Tahapan pengumpulan data dalam penelitian ini yaitu dengan mengambil data dari website SIPSN. Data yang digunakan merupakan data hasil capaian sampah di seluruh wilayah Indonesia beserta data hasil timbulan sampah dalam harian serta tahunan di seluruh wilayah Indonesia.

Provinsi	Kabupaten/Kota	Timbulan Sampah Tahunan (ton/tahun) (A)	Pengurangan Sampah Tahunan (ton/tahun) (B)	%Pengurangan Sampah(B/A)	Penanganan Sampah Tahunan (ton/tahun) (C)	%Penanganan Sampah(C/A)	Sampah Terkelola Tahunan (ton/tahun) (B+C)	%Sampah Terkelola(B+C)/A
Aceh	Kab. Aceh Selatan	35323.53	113.25	0.32	12775.00	36.17	12888.25	36.49
Aceh	Kab. Aceh Tenggara	41666.21	25.55	0.06	7227.00	17.34	7252.55	17.41
Aceh	Kab. Aceh Timur	65670.22	1044.80	1.59	22747.55	34.64	23792.35	36.23
Aceh	Kab. Aceh Tengah	40050.24	1093.18	2.73	22630.00	56.50	23723.18	59.23
Aceh	Kab. Aceh Barat	36813.72	3322.76	9.03	30502.76	82.86	33825.52	91.88

Gambar 2. Data sebelum diolah

Pada Gambar 2 merupakan tampilan data sebelum dilakukan tahapan *preprocessing*. Data tersebut ditampilkan dalam bentuk *dataframe* pada Google Colab.

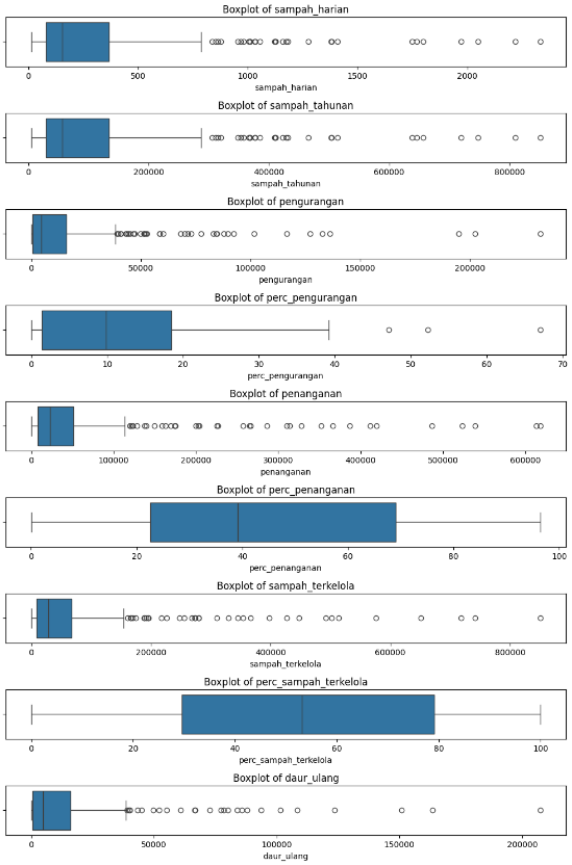
4.2. *Preprocessing*

Dalam penelitian ini, tahapan *Preprocessing* dilakukan untuk memastikan kualitas data sebelum digunakan dalam analisis *Clustering*. Untuk mempermudah analisis, data yang diperoleh digabungkan dengan tabel lain guna memperoleh informasi yang lebih lengkap. *Dataset* digabungkan dengan menggunakan Google Colab. Penggabungan dataset tersebut menggunakan *left join* dengan berdasarkan kolom Kabupaten/Kota. Hasil yang diperoleh adalah sebuah data *frame* yang terdiri dari 11 kolom dan 366 baris. Penjelasan terkait metadata tertera pada Tabel 1.

Tabel 1. Penjelasan metadata

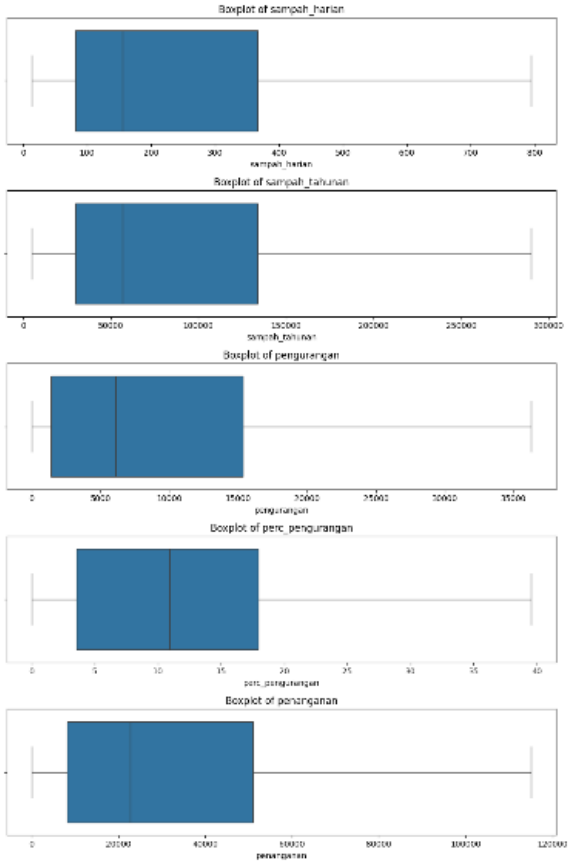
Variabel	Keterangan
Kabupaten/Kota	Nama Kabupaten/Kota
Sampah_harian	Total timbulan sampah dalam harian
Provinsi	Nama provinsi
Sampah_tahunan	Total timbulan sampah dalam setahun
Pengurangan	Pengurangan sampah
Perc_pengurangan	Persentase pengurangan sampah
Penanganan	Penanganan sampah
Perc_penanganan	Persentase penanganan sampah
Sampah_terkelola	Sampah yang terkelola
Daur_ulang	Sampah yang terdaur ulang

Selain itu, dilakukan tahapan deteksi *outlier* untuk mengidentifikasi nilai yang memiliki rentang terlalu jauh dari data lainnya. Dalam Gambar 2 merupakan representasi *outlier* pada data dengan menggunakan bantuan visualisasi *boxplot*.



Gambar 3. Data *outlier*

Sesuai dengan Gambar 3, didapatkan outlier pada variabel sampah_harian, sampah_tahunan, pengurangan, perc_pengurangan, penanganan, perc_penanganan, sampah_terkelola, perc_sampah terkelola, dan daur _ulang.



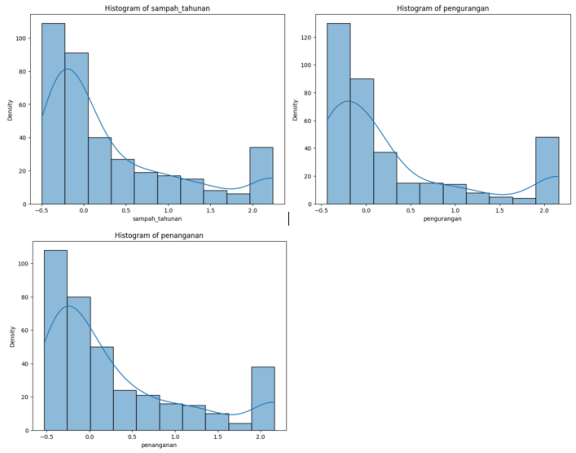
Gambar 4. Data *handling outlier*

Setelah didapatkan variabel-variabel yang mengalami outlier, dilanjutkan dengan penanganan outlier. Hasil dari penanganan outlier dapat dilihat dalam Gambar 4.

Setelah seluruh proses pembersihan data, dilakukan pemilihan variabel dan variabel yang dipilih adalah sampah_tahunan, pengurangan, dan penanganan. Ketiga variabel ini akan dilakukan ke dalam tahapan pemrosesan selanjutnya yaitu *feature scaling*. *Feature scaling* merupakan teknik untuk menyesuaikan nilai fitur dalam dataset agar berada dalam skala yang seragam atau dalam rentang tertentu. Langkah ini sangat penting untuk mencegah perbedaan skala yang dapat memengaruhi kinerja model *machine learning*, terutama pada algoritma yang sensitif terhadap ukuran data.

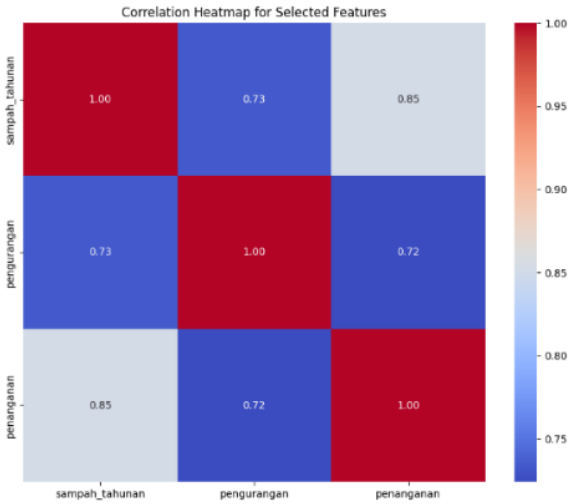
4.3. EDA

Pada tahapan eksplorasi data digunakan untuk mengetahui informasi dari data. Tahapan visualisasi dalam penelitian ini digunakan untuk melihat persebaran data pada tiap-tiap kolomnya. Persebaran tersebut menggunakan histogram.



Gambar 5. Histogram variabel

Histogram pada Gambar 5 menunjukkan distribusi berbagai variabel terkait pengelolaan sampah, di mana keseluruhan variabel memiliki pola distribusi *skewed* ke kanan.



Gambar 6. Korelasi antar variabel

Pada Gambar 6 merupakan heatmap korelasi, yang digunakan untuk menganalisis hubungan antar variabel dalam dataset pengelolaan sampah. Heatmap ini menampilkan koefisien korelasi Pearson, yang mengukur seberapa kuat hubungan antara dua variabel dalam skala -1 hingga 1. Berdasarkan Gambar 4 korelasi paling tinggi terdapat pada variabel sampah tahunan dan variabel penanganan dengan nilai 0.85 sedangkan korelasi paling rendah terdapat pada variabel penanganan dan variabel pengurangan yaitu 0.72.

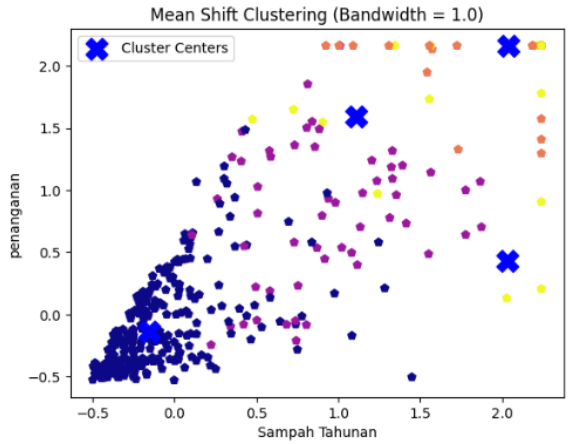
Tabel 2. Statistika Deskriptif

	Sampah	Pengurangan	Penanganan
Count	366.000000	366.000000	366.000000
Mean	0.330595	0.321800	0.310323
Std	0.816352	0.859076	0.827992
Min	-0.498376	-0.437713	-0.529575
25%	0.259896	-0.335672	-0.336426
50%	0.000000	0.000000	0.000000
75%	0.740104	0.664328	0.663574
Max	2.240104	2.164328	2.163574

Pada Tabel 1 menampilkan statistik deskriptif dari dataset yang berkaitan dengan pengelolaan sampah. Statistik ini mencakup jumlah data (*count*), rata-rata (*mean*), standar deviasi (std), nilai *minimum* (min), kuartil pertama (25%), median (50%), kuartil ketiga (75%), dan nilai maksimum (*max*) untuk setiap variabel.

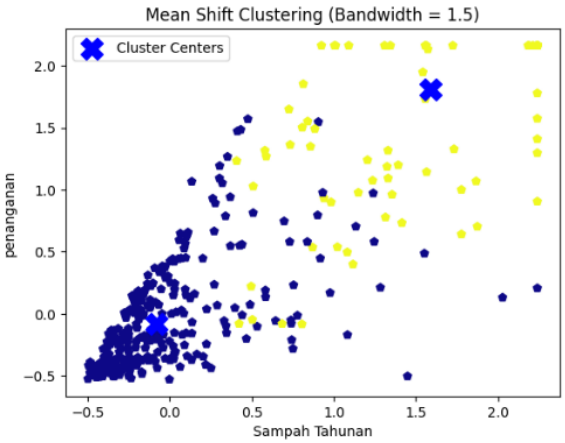
4.4. Pemodelan

Dalam penelitian ini pemodelan menggunakan algoritma *Mean Shift Clustering*. Dalam algoritma ini, tidak memerlukan penentuan nilai k untuk terbentuknya klaster. Akan tetapi algoritma ini menggunakan nilai *bandwidth* atau jangkauan untuk mendapatkan klaster.



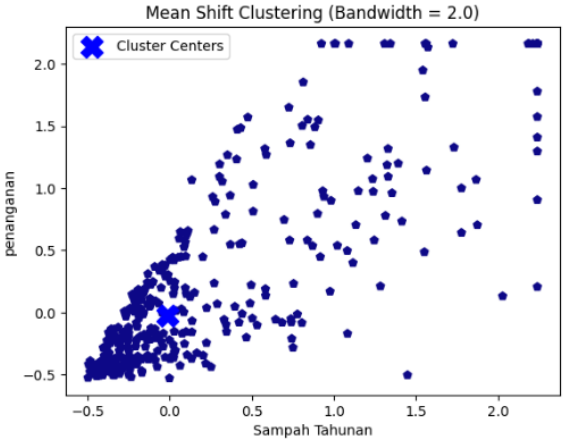
Gambar 7. Bandwidth 1.0

Pada Gambar 7 merupakan tampilan visualisasi dari algoritma dengan menggunakan *bandwidth* 1.0. Pada grafik, setiap titik mewakili suatu wilayah atau unit data, dengan warna berbeda menunjukkan keanggotaan dalam klaster yang berbeda. Sumbu horizontal merepresentasikan jumlah sampah yang dihasilkan dalam satu tahun, sedangkan sumbu vertikal menunjukkan tingkat penanganan sampah. Tanda X besar berwarna biru menandai pusat klaster yang diidentifikasi oleh algoritma. Dengan menggunakan nilai ini menghasilkan sebanyak 5 klaster.



Gambar 8. Bandwidth 1.5

Pada Gambar 8 merupakan tampilan visualisasi dari algoritma dengan menggunakan *bandwidth* 1.5. Setiap titik pada grafik merepresentasikan suatu wilayah atau unit data, di mana perbedaan warna menunjukkan klaster yang berbeda. Sumbu horizontal menggambarkan jumlah sampah yang dihasilkan dalam satu tahun, sedangkan sumbu vertikal menunjukkan tingkat penanganannya. Sementara itu, tanda X besar berwarna biru menandakan pusat klaster yang telah diidentifikasi oleh algoritma. Dengan menggunakan nilai ini menghasilkan sebanyak 2 klaster.



Gambar 9. Bandwidth 2.0

Pada Gambar 9 merupakan tampilan visualisasi dari algoritma dengan menggunakan *bandwidth* 1.0. Grafik ini menunjukkan titik-titik yang mewakili wilayah atau unit data, dengan variasi warna yang menandakan keanggotaan dalam klaster yang berbeda. Sumbu horizontal menunjukkan jumlah sampah yang dihasilkan dalam satu tahun, sedangkan sumbu vertikal menggambarkan tingkat penanganan sampah. Selain itu, tanda X besar berwarna biru menunjukkan pusat klaster yang telah ditentukan oleh algoritma. Dengan menggunakan nilai ini menghasilkan sebanyak 5 klaster.

4.5. Evaluasi

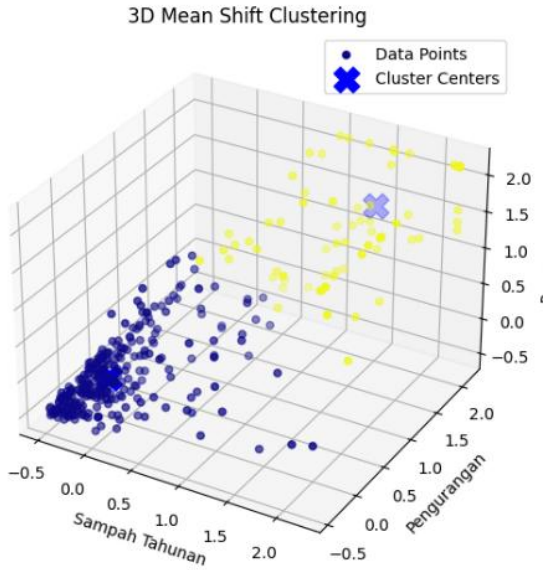
Untuk menentukan nilai bandwith yang tepat dapat menggunakan metrik evaluasi. Dalam *Clustering*, pada umumnya mengevaluasi model dengan menggunakan *silhouette score*.

Tabel 3. Perbandingan nilai *silhouette score*

Bandwidth	**Klaster**	***Silhouette score***
1.0	5	0.581
1.5	2	0.649
2.0	1	-1.0

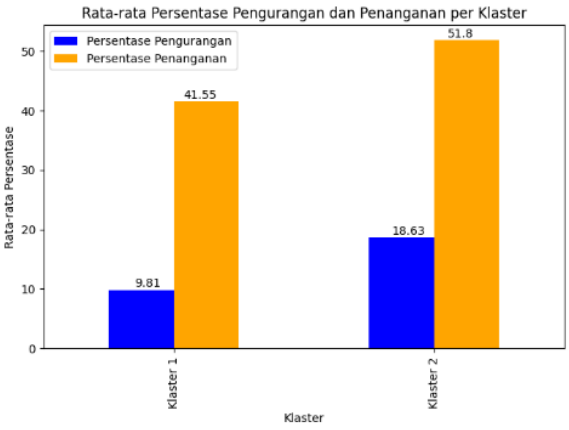
Hasil klasterisasi menggunakan *Mean Shift* dengan berbagai nilai *bandwidth* pada Tabel 1. Menunjukkan bahwa pemilihan *bandwidth* sangat mempengaruhi jumlah klaster dan kualitas pemisahan data. Pada *bandwidth* 1.0, terbentuk 5 klaster dengan *Silhouette Score* 0.581, yang menunjukkan pemisahan klaster cukup baik. Ketika *bandwidth* ditingkatkan menjadi 1.5, jumlah klaster berkurang menjadi 2, tetapi *Silhouette Score* meningkat menjadi 0.649, menandakan klaster yang lebih optimal dengan pemisahan lebih jelas. Namun, pada *bandwidth* 2.0, seluruh data tergabung dalam satu klaster dengan *Silhouette Score* -1.0, yang menunjukkan hasil klasterisasi tidak efektif. Dengan demikian, *bandwidth* 1.5 memberikan hasil terbaik karena memiliki keseimbangan antara jumlah klaster dan kualitas pemisahan data.

Berdasarkan nilai silhouette score yang didapat, bandwidth terbaik menggunakan 1.5



Gambar 10. Visualisasi Mean Shift

Pada gambar 10 merupakan hasil visualisasi 3 dimensi menggunakan algoritma mean shift dengan bandwidth 1.5. Terlihat bahwa tanda silang merupakan titik pusat masing-masing klaster sedangkan titik warna biru dan warna kuning merupakan sebaran data. Dalam visualisasi tersebut terlihat bahwa, penyebaran titik data pada kedua klaster terpisah dengan baik. Klaster 1 terdiri dari 284 data sedangkan klaster 2 terdiri dai 82 data. Dalam Gambar 11, menampilkan rata-rata persentase pengurangan dan penanganan pada masing-masing klaster. Pada klaster 1 didapatkan rata-rata persentase pengurangan sampah sebanyak 9.81% dengan persentase penanganan sampah sebanyak 41.55%. Pada klaster 2 didapatkan rata-rata persentase pengurangan sampah sebanyak 18.63% dengan persentase penanganan sampah sebanyak 51.8%. Dapat dikatakan bahwa jumlah sampah yang terkelola pada klaster 2 lebih tinggi dibandingkan dengan jumlah sampah yang terkelola pada klaster 1.



Gambar 11. Rata-rata klaster

Tabel 3 merupakan tabel wilayah beserta klasternya. Dalam tabel tersebut disebutkan masing-masing 5 wilayah pada tiap klaster .

Tabel 4. Wilayah Klaster

Wilayah	**Klaster**
Kab. Aceh Selatan	1
Kab. Aceh Tenggara	1
Kab. Aceh Timur	1
Kab. Aceh Tengah	1
Kab. Aceh Barat	1
Kab. Deli Serdang	2
Kota Medan	2
Kota Padang	2
Kota Pekanbaru	2
Kota Jambi	2

5. KESIMPULAN DAN SARAN

Hasil analisis *Clustering* menggunakan algoritma *Mean Shift* dengan *bandwidth* 1.5 menunjukkan bahwa metode ini mampu mengelompokkan data secara efektif dengan menghasilkan 2 klaster dan nilai silhouette sebesar 0.649. Klaster 1 dikategorikan dengan klaster tingkat pengelolaan sampah rendah sedangkan klaster 2 tingkat pengelolaan sampah tinggi. Nilai ini menunjukkan bahwa klaster yang terbentuk memiliki pemisahan yang cukup baik dan struktur yang jelas. Dengan demikian, penggunaan *bandwidth* 1.5 memberikan keseimbangan yang optimal antara jumlah klaster dan kualitas pemisahannya, menjadikannya pilihan yang tepat untuk segmentasi wilayah berdasarkan pengelolaan

sampah. Bagi penelitian selanjutnya, disarankan untuk membandingkan metode ini dengan algoritma *Clustering* lainnya guna mengevaluasi efektivitas segmentasi yang dihasilkan. Selain itu, eksplorasi terhadap berbagai nilai bandwidth juga perlu dilakukan untuk mengetahui apakah ada konfigurasi yang lebih optimal dalam menghasilkan jumlah dan kualitas klaster yang lebih baik. Dengan perbaikan dan pengembangan ini, diharapkan penelitian selanjutnya dapat memberikan wawasan yang lebih luas dan mendukung perumusan kebijakan pengelolaan sampah yang lebih efektif.

DAFTAR PUSTAKA

[1] T. Watiningsih, E. Sudaryanto, and D. Wahjudi, "Pemanfaatan Sampah Rumah Tangga Menjadi Kerajinan yang Lebih Bermanfaat," *WIKUACITYA J. Pengabdi. Kpd. Masy.*, vol. 3, no. 1, pp. 67–71, 2022, doi: 10.56681/wikuacitya.v3i1.145.

[2] A. Nagong, "Studi Tentang Pengelolaan Sampah Oleh Dinas Lingkungan Hidup Kota Samarinda Berdasarkan Peraturan Daerah Kota Samarinda Nomor 02 Tahun 2011 Tentang Pengelolaan Sampah," *J. Adm. Reform*, vol. 8, no. 2, p. 105, 2021, doi: 10.52239/jar.v8i2.4540.

[3] D. Syamsuar, "Analisis Interface dengan User Centered Design (UCD) pada Web SIPSN User Interface design analysis with User Centered Design (UCD) on the," vol. 5, pp. 1–6.

[4] SIPSN, "CAPAIAN KINERJA PENGELOLAAN SAMPAH," Sistem Informasi Pengelolaan Sampah Nasional. [Online]. Available: https://sipsn.menlhk.go.id/sipsn/public/data/capaian

[5] BPK RI, "Undang-undang (UU) Nomor 18 Tahun 2008 tentang Pengelolaan Sampah," Badan Pemeriksa Keuangan Republik Indonesia. Accessed: Mar. 10, 2025. [Online]. Available: https://peraturan.bpk.go.id/Details/39067/uu-no-18-tahun-2008

[6] M. Rafly, H. Putra, and P. Priyana, "Upaya Penanggulangan Tempat Penampungan Sementara Di Dusun Kaum Jaya Serta Dampak Yang Timbul Bagi Lingkungan Dan Masyarakat," *JUSTITIA J. Ilmu Huk. dan Hum.*, vol. 9, no. 2, pp. 898–915, 2022.

[7] W. Wahyu Pribadi, A. Yunus, and A. S. Wiguna, "Perbandingan Metode K-Means Euclidean Distance Dan Manhattan Distance Pada Penentuan Zonasi Covid-19 Di Kabupaten Malang," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 493–500, 2022, doi: 10.36040/jati.v6i2.4808.

[8] R. Rizuan, E. Haerani, J. Jasril, and L. Oktavia, "Penerapan Algoritma Mean-Shift Pada Clustering Penerimaan Bantuan Pangan Non Tunai," *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 1019–1027, 2023, doi: 10.47065/josyc.v4i4.3876.

[9] T. A. Cinderatama, R. Z. Alhamri, and Y. Yunhasnawa, "Implementasi Metode K-Means, Dbscan, Dan Meanshift Untuk Analisis Jenis Ancaman Jaringan Pada Intrusion Detection System," *INOVTEK Polbeng - Seri Inform.*, vol. 7, no. 1, p. 169, 2022, doi: 10.35314/isi.v7i1.2336.

[10] F. Salsabila, T. Ridwan, and H. H, "Analisa Volume Penyebaran Sampah Di Karawang Menggunakan Algoritma K-Means Clustering," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, 2024, doi: 10.23960/jitet.v12i2.4226.

[11] V. E. Lumbantobing, Laili Fitria, and H. Sutrisno, "Analisis Potensi Nilai Ekonomi Sampah Plastik," *J. Alwatzikhoebillah Kaji. Islam. Pendidikan, Ekon. Hum.*, vol. 9, no. 1, pp. 251–262, 2023, doi: 10.37567/alwatzikhoebillah.v9i1.1663.

[12] I. S. Abidin and D. S. H. Marpaung, "Observasi Penanganan dan Pengurangan Sampah di Universitas Singaperbangsa Karawang," *JUSTITIA J. Ilmu Huk. dan Hum.*, vol. 8, no. 4, pp. 872–882, 2021.

[13] R. F. Utari, F. Insani, S. Agustian, and L. Afriyanti, "Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Mean Shift," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, pp. 1015–1023, 2024, doi: 10.57152/malcom.v4i3.1428.

[14] H. Alexander, Y. Umaidah, and M. Jajuli, "Implementasi Clustering Untuk Menentukan Efektivitas Nilai Siswa Sesudah Pandemi Covid-19 Menggunakan Algoritma K-Means," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 3, pp. 1493–1500, 2023, doi: 10.36040/jati.v7i3.7174.

[15] A. A. Harahap, M. Raihan, N. Amani, and R. Andini, "Perbandingan Teknik Unsupervised Learning untuk Pengelompokan Data Jumlah Desa Di Indonesia," *SENTIMAS Semin. Nas. Penelit. dan Pengabdi. Masy.* , pp. 163–170, 2023, [Online]. Available: https://journal.irpi.or.id/index.php/sentimas

[16] M. A. Khowarizmi, "Algoritma Mean Shift untuk Menentukan Segmentasi Pelanggan pada Penjualan Toko Online," vol. 3, pp. 1–7, 2021.

[17] A. Luthfia, H. Yahman, J. Sianturi, F. Zaharani, U. N. Medan, and M. S. Clustering, "PENERAPAN ALGORITMA MEAN SHIFT CLUSTERING UNTUK SEGMENTASI PELANGGAN WHOLESALE BERDASARKAN POLA," vol. 10, no. 11, pp. 44–55, 2024.

[18] Y. Nurohmah, R. Mayasari, and B. Nurina Sari, "Optimalisasi Performa K-Means Clustering Dengan Pca Dalam Analisis Tingkat Kemiskinan Di Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 3, pp. 1657–1665, 2023, doi: 10.36040/jati.v7i3.6884.

[19] T. M. Fahrudin, P. A. Riyantoko, K. M. Hindrayani, and M. H. P. Swari, "Cluster Analysis of Hospital Inpatient Service Efficiency Based on BOR, BTO, TOI, AvLOS Indicators using Agglomerative Hierarchical Clustering," *Telematika*, vol. 18, no. 2, p. 194, 2021, doi: 10.31315/telematika.v18i2.4786.

[20] P. R. Harnanda, N. Damastuti, and T. M. Fahrudin, "GIS implementation and classterization of potential blood donors using the agglomerative hierarchical clustering method," *IJEEIT Int. J. Electr. Eng. Inf. Technol.*, vol. 3, no. 2, pp. 44–54, 2021, doi: 10.29138/ijeeit.v3i2.1305.