



## **Clustering Daerah Penyumbang Sampah Berdasarkan Provinsi di Indonesia Menggunakan Algoritma *K-Means***

**Desi Kristina Sitinjak<sup>1</sup>, Betha Nurina Sari<sup>2</sup>, Iqbal Maulana<sup>3</sup>**

Universitas Singaperbangsa Karawang

### **Abstract**

Received: 7 Agustus 2022  
Revised: 10 Agustus 2022  
Accepted: 16 Agustus 2022

*The problem of waste is still one of the big problems that occur in Indonesia today. Along with the development of rapid population growth, it directly contributes negatively which results in wider and denser residential areas as well as a sharp increase in waste production. Many people in Indonesia are not able to keep the environment clean, so the current waste problem cannot be handled properly. The losses obtained from this waste problem must be minimized. The step that can be taken is to determine the area of waste contributors in Indonesia as an initial effort in dealing with waste problems in Indonesia. In this study, clustering of waste contributor areas based on provinces in Indonesia was carried out using the *K-Means* algorithm and mapping was carried out using QGIS. Clusters are divided into 2, namely clusters of areas that contribute to high and low levels of waste. The results of grouping waste contributor areas based on provinces in Indonesia using the *K-Means* algorithm were found that 30 provinces were low clusters, and 4 provinces were high clusters. The modeling results were evaluated using the Davies Bouldin Index to determine the quality of the cluster. The results of the cluster test obtained an index value of 0.329 for  $k = 2$ , this is the cluster with the best DBI value. This study is expected to provide a presentation of clustering data on waste contributor areas based on provinces in Indonesia which can then be used to assist the government in improving the quality of the solid waste management system.*

**Keywords:** *Algoritma K-Means, Clustering, Data Mining, QGIS, Waste*

(\* Corresponding Author: [desikristina102@gmail.com](mailto:desikristina102@gmail.com))

**How to Cite:** Kristina, D., Sari, B., & Maulana, I. (2022). Clustering Daerah Penyumbang Sampah Berdasarkan Provinsi di Indonesia Menggunakan Algoritma *K-Means*. *Jurnal Ilmiah Wahana Pendidikan*, 8(16), 137-146. <https://doi.org/10.5281/zenodo.7059032>

## **PENDAHULUAN**

Permasalahan sampah masih menjadi salah satu permasalahan besar yang terjadi di negara Indonesia saat ini. Sudah semestinya permasalahan sampah ini menjadi tanggung jawab semua warga negara dan pemerintah. Menurut *World Health Organization* (WHO) sampah adalah sesuatu yang tidak terpakai, tidak disukai atau dibuang begitu saja, yang disebabkan oleh aktivitas manusia dan tidak terjadi begitu saja (Chandra, 2006). Seiring berkembangnya pertumbuhan penduduk yang pesat secara langsung memberikan kontribusi negatif yang mengakibatkan daerah pemukiman semakin luas dan padat juga peningkatan tajam terhadap produksi sampah (Yanto, 2018).

Indonesia merupakan negara penyumbang sampah terbesar kedua di dunia. Banyak masyarakat di Indonesia yang tidak mampu menjaga kebersihan

lingkungan, sehingga masalah sampah saat ini belum bisa ditangani dengan baik (Silitonga et al., 2019).

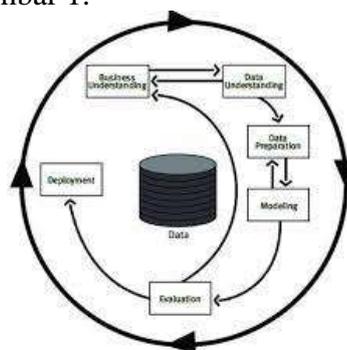
Penumpukan sampah akibat pencemaran penduduk di daerah terbuka menyebabkan pencemaran tanah yang berdampak pada saluran air tanah. Demikian juga masalah pembakaran sampah yang tidak kalah serius dampaknya dikarenakan sampah yang dibakar akan menghasilkan zat atau gas yang dapat menyebabkan gangguan kesehatan pemicu kanker (karsinogenik), bahkan kematian (Puriana et al., 2021).

Kerugian yang didapat dari permasalahan sampah ini tentunya harus diminimalisir. Salah satu langkah yang dapat dilakukan adalah dengan menentukan daerah penyumbang sampah di Indonesia sebagai upaya awal dalam menangani permasalahan sampah di Indonesia. Oleh karena itu diperlukan proses *data mining* untuk melakukan pengelompokan provinsi penyumbang sampah di Indonesia. Penelitian ini nantinya akan menerapkan salah satu teknik *unsupervised learning* yaitu *clustering*. *Clustering* merupakan aktivitas (*task*) yang bertujuan mengelompokkan data yang menunjukkan kemiripan antara dua data dikelompokkan ke dalam *cluster* atau kelompok dimana data dalam suatu *cluster* memiliki tingkat kemiripan (*similarity*) paling tinggi dan data antar *cluster* merupakan aktivitas (tugas) yang paling sedikit memiliki kemiripan (Nishom, 2019).

Penelitian ini menerapkan metode *K-Means clustering* yang merupakan metode sederhana dan memiliki kemampuan dalam mengelompokkan data dengan baik dan dalam jumlah yang cukup besar dengan waktu komputasi yang cepat dan efisien. Terbukti dalam penelitian Artanto et al., (2019) bahwa implementasi algoritma *K-Means* lebih baik daripada implementasi *Fuzzy C-Means*. Maka pada penelitian ini metode yang digunakan adalah *clustering* dengan algoritma *K-Means* untuk pengelompokan daerah penyumbang sampah di Indonesia berdasarkan provinsi. Dalam memvisualisasikan hasil pemetaan dari penelitian ini akan dilakukan dengan bantuan *tools* QGIS.

## METODOLOGI PENELITIAN

Metodologi yang digunakan pada penelitian ini adalah metodologi CRISP-DM (*Cross Industry Standard Process for Data mining*) memiliki enam fase siklus hidup yang dapat dilihat pada Gambar 1.



**Gambar 1** Tahapan Metodologi CRISP-DM  
(Sumber: Chapman, 2000:10)

a. *Business Understanding*

Tahap ini yaitu tahap pemahaman bisnis tentang penggunaan data yang diimplementasikan berdasarkan kebutuhan bisnis, dengan berfokus pada pemahaman tujuan dan kebutuhan dari perspektif bisnis yang kemudian mentransfer pengetahuan itu untuk mengidentifikasi masalah internal. Selain itu, rencana dan strategi dikembangkan untuk mencapai tujuan tersebut. Selanjutnya akan ditentukan rencana dan strategi untuk mencapai tujuan tersebut.

b. *Data Understanding*

*Data Understanding* yaitu fase pemahaman dari data yang dimiliki, pada tahap ini dilakukan pengumpulan dan pemahaman terhadap data awal yang akan digunakan, mengeksplorasi data untuk mendapatkan *insight* pada data, dan mengidentifikasi masalah yang berkaitan dengan kualitas data.

c. *Data Preparation*

Fase *data preparation* yaitu fase yang mencakup pemilihan tabel, *record*, dan atribut-atribut data, termasuk proses pembersihan dan transformasi data untuk kemudian dijadikan masukan dalam tahap pemodelan (*modeling*). Pada fase ini dilakukannya variabel yang akan dianalisis atau fase pengolahan data, meliputi kegiatan-kegiatan yang digunakan untuk memproses *dataset* akhir (*data* yang akan diproses pada tahap *modeling*) dari data mentah. Fase ini dapat diulang hingga beberapa kali.

d. *Modeling*

Fase ini yaitu fase pemodelan menentukan dan penerapan teknik dari *data mining* dan algoritma yang akan digunakan dengan menentukan parameter dengan nilai yang optimal. Pada fase ini, berbagai teknik pemodelan dipilih dan diterapkan dari beberapa parameter disesuaikan dengan nilai optimal. Secara khusus, ada beberapa teknik berbeda yang dapat diterapkan untuk masalah *data mining* yang sama. Pada fase ini masih memungkinkan kembali ke tahap sebelumnya.

e. *Evaluation*

Fase *evaluation* yaitu tahap evaluasi dari penerapan model yang digunakan apakah sudah telah sesuai dengan tujuan atau tidak. Pada fase ini model sudah terbentuk lebih baik dalam hal menganalisis data yang diharapkan. Sementara itu, pada fase ini dilakukan validasi model dan menentukan apakah model tersebut dapat memenuhi tujuan yang ditetapkan pada fase pertama (*Business Understanding*). Kunci pada fase ini adalah menentukan apakah ada masalah bisnis yang belum terselesaikan. Di akhir langkah ini, harus ditentukan penggunaan hasil dari proses *data mining*.

f. *Deployment*

Fase *deployment* merupakan fase penyebaran atau presentasi hasil pengetahuan yang didapatkan dari proses *data mining*. Pada tahap ini juga pengetahuan atau informasi yang diperoleh disimpan dan disampaikan dalam bentuk khusus sehingga pengguna dapat menggunakannya. Fase *deployment* dapat berupa pembuatan laporan sederhana atau mengimplementasikan proses *data mining* yang berulang dalam sebuah unit bisnis.

## KAJIAN TEORI

### *Data mining*

*Data mining* adalah istilah yang menggambarkan perolehan pengetahuan dalam *database* (Buulolo, 2020). Pada penelitian yang dilakukan oleh Purba et al., (2019) *Data mining* merupakan proses menemukan korelasi baru yang bermanfaat, sampel dengan menambang sejumlah besar gudang data, menggunakan teknologi pengenalan pola seperti statistik dan metode matematika. Dari beberapa pemaparan tentang *data mining* maka dapat dikatakan bahwa *data mining* merupakan sebuah proses mengolah suatu data menjadi informasi.

### *Clustering*

Menurut Nishom (2019) *clustering* merupakan tindakan yang dirancang untuk mengelompokkan data ketika ada kesamaan antara data dalam cluster. Berdasarkan penelitian Nabila et al., (2021) Cluster adalah kelompok atau kumpulan objek data yang serupa. *Clustering* merupakan salah satu metode data mining dimana objek-objek tersebut dibagi menjadi beberapa bagian. Jadi dapat disimpulkan bahwa *clustering* adalah metode untuk pengelompokan data yang memiliki kesamaan atau kemiripan sehingga dapat dilakukan pelabelan sesuai keinginan.

### *Algoritma K-Means*

*Algoritma K-Means* adalah salah satu algoritma *clustering* yang masuk ke dalam kelompok *unsupervised learning* dan digunakan untuk mengelompokkan data ke dalam beberapa kelompok dengan sistem partisi (Wanto et al., 2020). Menurut Wanto et al., (2020) dalam bukunya yang berjudul “*Data mining : Algoritma dan Implementasi*” mengatakan langkah-langkah algoritma *K-Means* adalah sebagai berikut :

1. Tentukan nilai k atau jumlah *cluster* pada data set.
2. Menentukan nilai pusat (*centroid*).
3. Menghitung jarak terhadap masing-masing *centroid* menggunakan rumus jarak *euclidean distance* sehingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Berikut persamaan 1 rumus *euclidean distance*:

$$d(x, y) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2}$$

Dimana :

d(x,y) = jarak antara data pada titik x dan y

x = Titik data objek

y = titik data *centroid*

i = jumlah atribut data

4. Melakukan klasifikasi data berdasarkan kedekatannya dengan *centroid* (jarak terkecil).

5. Memperbaharui nilai *centroid*. Nilai *centroid* ini baru diperoleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan persamaan 2 berikut:

$$C_k = \left(\frac{1}{n_k}\right) \sum d_i$$

Dimana :

C<sub>k</sub> = *centroid cluster*

n<sub>k</sub> = jumlah data dalam *cluster* K

$d_i$  = jumlah dari nilai jarak yang masuk dalam masing masing *cluster*

6. Ulangi langkah nomor 2-5 apabila posisi centroid baru tidak sama.

### ***Davies Bouldin Index***

Pada penelitian yang dilakukan oleh Jumadi Dehotman Sitompul et al., (2019) menyatakan bahwa DBI atau *Davies Bouldin Index* merupakan salah satu metode yang digunakan untuk mengukur validitas *cluster* pada suatu metode pengelompokan. Pada penelitian yang dilakukan oleh Radiyanto et al., (2022) didapatkan bahwa jika perhitungan nilai Davies Bouldin Index (DBI) mendekati nilai 0, maka itu adalah nilai *cluster* yang paling baik.

## **HASIL & PEMBAHASAN**

Hasil dari penelitian yang telah dilakukan adalah bagaimana melakukan implementasi algoritma *K-Means clustering* untuk menghasilkan pengelompokan daerah penyumbang sampah berdasarkan provinsi di Indonesia yang selanjutnya dievaluasi dan divisualisasikan menggunakan *software QGIS*.

### ***a. Business Understanding***

Pada langkah ini, dilakukan analisis masalah yaitu permasalahan sampah dengan kerugiannya yang harus diminimalisir. Maka dari itu penelitian ini melakukan proses *data mining* untuk menghasilkan *clustering* daerah penyumbang sampah berdasarkan provinsi di Indonesia sebagai langkah awal dalam menanggulangi permasalahan sampah dengan menentukan daerah penyumbang sampah berdasarkan provinsi di Indonesia.

### ***b. Data Understanding***

#### ***1. Collect Initial Data***

Data yang digunakan meliputi data sampah masuk, sampah organik terolah, sampah anorganik terolah, *recovery* pemulung dan jumlah penduduk pada tahun 2020 dan data tersebut bersumber dari website Kementerian Lingkungan Hidup dan Kehutanan (<https://sipsn.menlhk.go.id/>) dan dari *website* Badan Pusat Statistik (<http://bps.go.id/>).

#### ***2. Describe Data***

Tahap mendeskripsikan data dilakukan setelah proses pengumpulan data selesai, tahap ini bertujuan untuk memahami data yang akan diolah dalam pengelompokan daerah penyumbang sampah berdasarkan provinsi di Indonesia.

#### ***3. Explore Data***

Tahap ketiga dalam *Data Understanding* adalah eksplorasi terhadap isi data. Dari data yang ada berupa data dalam bentuk tabel *excel* untuk setiap atribut. Tahap eksplorasi data berfungsi untuk mengetahui karakteristik data yang digunakan.

#### ***4. Verify Data Quality***

Tahapan memverifikasi kualitas data merupakan tahapan menilai kualitas *dataset* yang dimiliki dari kelengkapan dan kesesuaian data, memeriksa data dari nilai-nilai yang hilang atau kosong, memeriksa apakah atributnya sudah lengkap atau tidak. *Dataset* berupa rekap data sampah masuk, sampah organik terolah, sampah anorganik terolah, *recovery* pemulung dan jumlah penduduk. Nilai nol tersebut perlu diatasi untuk meningkatkan kualitas data.

**c. Data Preparation**

Pada tahap *data preparation* dilakukan aktivitas pengolahan data awal yang bertujuan untuk membuat *dataset* final. Dalam penelitian ini, data daerah penyumbang sampah berdasarkan provinsi di Indonesia akan dibersihkan terlebih dahulu dari atribut data yang tidak diperlukan saat proses pemodelan dilakukan. Kegiatan persiapan data dengan membangun suatu *dataset* untuk menyesuaikan dengan kebutuhan pada pemodelan meliputi semua kegiatan untuk membangun *dataset* yang akan dimasukkan ke dalam alat pemodelan *data mining* dari data mentah awal.

**1. Select Data**

Data diambil tentang daerah penyumbang sampah berdasarkan provinsi di Indonesia. Data yang digunakan meliputi data sampah masuk, sampah organik terolah, sampah anorganik terolah, *recovery* pemulung dan jumlah penduduk pada tahun 2020 disimpan dalam bentuk file excel atau \*.xlsx atau \*.xls. Penggalan *dataset* yang telah diseleksi dapat dilihat pada Tabel 1.

**Tabel 1** Data Daerah Penyumbang Sampah di Indonesia

Provinsi	Sampahmasuk (ton/thn)	Sampahmasuk Landfill (ton/thn)	Sampah Organikterolah (ton/thn)	Sampah An-Organikterolah (ton/thn)	RecoveryPemulung (ton/thn)
Aceh	13,034.15		11,391.65	0.00	0.00
Aceh	3,854.40		3,671.90	0.00	182.50
Aceh	5,420.25		5,237.75	0.00	182.50
Aceh	22,229.81		22,229.81	0.00	0.00
Aceh	481.80		481.80	0.00	0.00
Aceh	4,818.00		4,818.00	0.00	0.00
Aceh	6,245.15		5,763.35	0.00	299.30
Aceh	9,125.00		6,132.00	73.00	2,920.00
Aceh	21,900.00		21,535.00	0.00	365.00
Aceh	67,633.55		67,633.55	0.00	0.00
Aceh	6,094.88		4,561.65	1,533.23	0.00
Aceh	6,380.20		6,015.20	0.00	365.00
Aceh	17,520.00		16,775.40	730.00	14.60
Sumatera Utara	3,285.00		3,285.00	0.00	0.00
Sumatera Utara	1,095.00		1,095.00	0.00	0.00
Sumatera Utara	5,475.00		5,475.00	0.00	0.00
Sumatera Utara	16,425.00		15,549.00	730.00	146.00
Sumatera Utara	149,650.00		138,700.00	3,650.00	1,825.00
Sumatera Utara	0.00		0.00	0.00	0.00
Sumatera Utara	157,909.95		157,362.45	0.00	547.50
Sumatera Utara	10,238.25		10,055.75	0.00	182.50
Sumatera Utara	11,178.13		11,041.25	91.25	45.63
Sumatera Utara	657.00		642.40	0.00	14.60
Sumatera Utara	9,672.50		8,942.50	0.00	730.00
Sumatera Utara	109.50		109.50	0.00	0.00
Sumatera Utara	328,500.00		224,475.00	91,250.00	0.00
Sumatera Utara	61,685.00		61,137.50	182.50	365.00
Sumatera Utara	20,440.00		15,038.00	730.00	2,080.50
					2,591.50

**2. Data Cleaning**

Pada tahap pembersihan data (*data cleaning*) seluruh atribut yang ada pada *dataset* diseleksi apakah atribut tersebut terdapat *missing value* atau tidak sehingga mendapatkan atribut-atribut yang berisi nilai yang relevan. *Missing value*, merupakan syarat yang harus dikerjakan dalam *data mining* sehingga akan diperoleh *dataset* yang bersih untuk digunakan pada tahap *modeling*. Data dikatakan *missing value* bila atribut-atribut dalam *dataset* tidak berisi nilai atau kosong. Pada Tabel 2 merupakan contoh penggalan *dataset* sebelum *cleaning*.

**Tabel 2** *Dataset* sebelum *cleaning*

Row No.	Provinsi	Sampah ma...	Sampah Org...	Sampah An...	Recovery Pe...	Jumlah Pen...
1	Aceh	184737.190	2336.230	299.300	5854.600	5274.900
2	Sumatera Ut...	807647.490	99944.730	3951.130	22627.100	14799.400
3	Sumatera Ba...	385943.050	1271.190	171.650	22715.050	5534.500
4	Riau	539122.900	2551.230	5588.190	23496.380	6394.100
5	Jambi	202274.140	3036.430	1825	8716.200	3548.200
6	Sumatera Sel...	577559.850	2626.910	15389.860	11866.150	8467.400
7	Bengkulu	40814.300	766.500	2091.090	1022	2010.700
8	Lampung	266888	9033.750	?	10194.450	9007.800
9	Kepulauan B...	102812.260	2019.180	273.750	1058.500	1455.700
10	Kepulauan Ri...	487156.350	7219.700	684.740	7936.550	2064.500
11	DKI Jakarta	2295488.980	?	?	?	10562.100
12	Jawa Barat	1683754.920	9875.080	5947.070	41157.400	48274.200
13	Jawa Tengah	1498819.110	295589.300	69003.250	82891.180	36516
14	D.I. Yogyakarta	248449.380	2528	365	7187.530	3668.700

**Format Data**

Tahapan ini adalah memproduksi set data akhir yang siap ditambah atau diolah dalam *tools* pemodelan *data mining* kemudian disimpan ke dalam bentuk *.xlsx* diubah menjadi format *CSV (Comma Separated Values)* atau *.csv*. Hal itu untuk mempermudah proses *modeling* yang akan dilakukan pada *tools Rapid miner*.

**d. Modeling**

*Modeling* adalah fase yang secara langsung melibatkan teknik *data mining*. Dalam proses pembangunan model algoritma *K-Means* akan dihitung dengan menggunakan rumus *Euclidean Distance* dengan menggunakan *tools Rapid miner*. Pada saat melakukan proses *K-Means Clustering*, *k cluster* yang digunakan harus ditentukan dahulu untuk mendapatkan hasil yang terbaik. Proses metode *clustering* dengan menggunakan algoritma *K-Means* akan dilakukan terhadap data daerah penyumbang sampah di Indonesia.

**Cluster Model**

Cluster 0: 30 items  
 Cluster 1: 4 items  
 Total number of items: 34

**Gambar 2** *Cluster Model*

**Tabel 3** Anggota *Cluster*

<i>Cluster</i>	Anggota	Jumlah	Kriteria
<i>Cluster 1</i>	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, D.I Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Timur, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Utara, Sulawesi Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua, Papua Barat	30	Rendah
<i>Cluster 2</i>	DKI Jakarta, Jawa Barat, Jawa tengah, Jawa Timur	4	Tinggi

**e. Evaluasi**

Setelah dilakukan analisis pada *cluster* dengan menggunakan algoritma *K-Means* tahap selanjutnya dilakukan evaluasi terhadap model jumlah *cluster* yang paling optimal dengan menggunakan *Davies Bouldin Index* (DBI).

**Tabel 4** Hasil *Davies Bouldin Index*

No	Jumlah <i>K Cluster</i>	Hasil <i>Davies Bouldin Index</i>
1	K = 2	0.329
2	K = 3	0.427
3	K = 4	0.422
4	K = 5	0.373

Setelah dilakukannya perbandingan ternyata didapatkan hasil bahwa jumlah *cluster* yang baik adalah k = 2. Hasil tersebut dikatakan baik karena, pada k *cluster* sebanyak 2 didapatkan nilai DBI yang mendekati 0 dibandingkan k *cluster* 3,4 dan 5.

**f. Deployment**

Laporan akhir mengenai pengetahuan yang akan didapat atau pengenalan pada data dalam proses *data mining* dan dipresentasikan dalam bentuk deskripsi yang mudah dipahami.



**Gambar 3** Peta *cluster* daerah penyumbang sampah di Indonesia 2020

*Cluster* daerah penyumbang sampah di Indonesia berdasarkan provinsi tahun 2020 terbagi menjadi 2 *cluster*. Daerah dengan warna hijau merupakan *cluster 1* yang berarti daerah penyumbang sampah tingkat rendah, sedangkan daerah dengan warna merah merupakan *cluster 2* yang berarti daerah penyumbang sampah tingkat tinggi.

## KESIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan maka kesimpulan yang didapatkan adalah sebagai berikut:

1. Penerapan algoritma *K-Means* mampu mengetahui pengelompokan daerah penyumbang sampah di Indonesia berdasarkan sampah masuk, sampah organik terolah, sampah anorganik terolah, *recovery* pemulung dan jumlah penduduk. Pada pengelompokan daerah penyumbang sampah berdasarkan provinsi di Indonesia didapatkan sebanyak 30 provinsi masuk ke dalam *cluster* tingkat rendah dan 4 provinsi masuk ke dalam *cluster* tingkat tinggi.
2. Hasil *clustering* yang didapatkan menggunakan algoritma *K-Means* dengan bantuan evaluasi menggunakan *Davies Bouldin Index* (DBI) menghasilkan jumlah *cluster* sebanyak 2 *cluster*. Penentuan kualitas jumlah nilai DBI adalah semakin rendah nilai DBI mendekati nilai 0, maka itu adalah nilai *cluster* yang paling baik. Pada penelitian yang telah dilakukan nilai DBI adalah 0.329.
3. Hasil visualisasi daerah penyumbang sampah di Indonesia berdasarkan provinsi pada tahun 2020 melalui Quantum GIS didapatkan daerah dengan warna hijau merupakan *cluster* 1 yang berarti daerah penyumbang sampah tingkat rendah, sedangkan daerah dengan warna merah merupakan *cluster* 2 yang berarti daerah penyumbang sampah tingkat tinggi.

## SARAN

Dari hasil penelitian yang telah dilakukan, maka saran-saran diberikan untuk menjadi referensi dalam penelitian selanjutnya, yaitu:

1. Diharapkan untuk penelitian selanjutnya dapat menambahkan *dataset* yang lebih banyak, karena semakin banyak data semakin berpengaruh untuk mendekati tujuan yang ingin dicapai.
2. Disarankan untuk menggunakan algoritma atau metode *clustering* yang lain dalam melakukan proses *clusterisasi*.
3. Disarankan untuk menggunakan tools bantuan *data mining* yang lain, yang lebih memungkinkan mempunyai kemampuan yang lebih tepat dan akurat dalam melakukan analisis pengelompokan.
4. Disarankan juga untuk menggunakan parameter evaluasi hasil *clustering* terhadap model jumlah *cluster* selain *Davies Bouldin Index* (DBI).

## DAFTAR PUSTAKA

- Artanto, H., Istiadi, Marisa, F., & Purnomo, D. (2019). Implementasi Dan Komparasi Algoritma Fuzzy C-Means Dan K-Means Untuk Mengelompokkan Siswa Berdasarkan Nilai Akademik Dan Perilaku Siswa ( Data Survey ). *Conference on Innovation and Application of Science and Technology (CIASTECH 2019)*, 287–292.
- Buulolo, E. (2020). *Data Mining Untuk Perguruan Tinggi*. Yogyakarta: Deepublish.
- Chandra, B. (2006). *Pengantar Kesehatan Lingkungan*. Jakarta: EGC.
- Jumadi Dehotman Sitompul, B., Salim Sitompul, O., & Sihombing, P. (2019). Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm. *Journal of Physics: Conference Series*, 1235(1), 6–12. <https://doi.org/10.1088/1742->

6596/1235/1/012015

- Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(2), 100. <http://jim.teknokrat.ac.id/index.php/JTISI>
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 20–24. <https://doi.org/10.30591/jpit.v4i1.1253>
- Purba, W., Siawin, W., & . H. (2019). Implementasi Data Mining Untuk Pengelompokan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan Algoritma K-Means Clustering. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, 2(2), 85–90. <https://doi.org/10.34012/jusikom.v2i2.429>
- Puriana, R. H., Mardhika, R., Mulyono, Faruq, M. M., & Suharti, Abd. Cholid, H. (2021). Meningkatkan Kesadaran Masyarakat Untuk Menjaga Kebersihan Lingkungan Dengan Cara Membuang Sampah Pada Tempatnya Dan Cara Pengelolaan Sampah. *Kanigara*, 1(2), 173–178.
- Radiyanto, D., R, U. A., & Muharom, L. A. (2022). Clustering Untuk Menentukan Strategi Promosi Universitas Muhammadiyah Jember Dengan Algoritma K-Medoids Clustering To Determine The University Of Muhammadiyah Jember Promotion Strategy With K-Medoids Algorithm. *Jurnal Smart Teknologi*, 3(3), 312–320.
- Silitonga, D. A., Windarto, A. P., & Hartama, D. (2019). Penerapan Metode K-Medoid pada Pengelompokan Rumah Tangga Dalam Perlakuan Memilah Sampah Menurut Provinsi. *Seminar Nasional Sains & Teknologi Informasi (SENSASI) SENSASI 2019 ISBN:*, 313–318.
- Wanto, A., Siregar, M. N. H., Windarto, A. P., Hartama, D., Ginantra, N. L. W. S. R., Napitupulu, D., Negara, E. S., Lubis, M. R., Dewi, S. V., & Limbong, C. P. T. (2020). *Data Mining: Algoritma dan Implementasi*. Medan: Yayasan bKita Menulis.
- Yanto, R. (2018). Implementasi Data Mining Estimasi Ketersediaan Lahan Pembuangan Sampah menggunakan Algoritma Simple Linear Regression. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 2(1), 361–366.