



Tersedia online di [www.journal.unipdu.ac.id](http://www.journal.unipdu.ac.id)  
Unipdu  
Terakreditasi Sinta S5

Halaman jurnal di [www.journal.unipdu.ac.id/index.php/teknologi](http://www.journal.unipdu.ac.id/index.php/teknologi)



## Penerapan metode *k-means clustering* data COVID-19 di Provinsi Jakarta

### Implementation of *k-means method* for COVID-19 data clustering in Jakarta Province

Meida Cahyo Untoro <sup>a</sup>, Leslie Anggraini <sup>b,\*</sup>, Maria Andini <sup>c</sup>, Hesti Retnosari <sup>d</sup>, M. Anas Nasrulloh <sup>e</sup>

<sup>a,b,c,d,e</sup> Teknik Informatika, Institut Teknologi Sumatera, Lampung Selatan, Indonesia

email: <sup>a</sup> [cahyo.untoro@if.its.ac.id](mailto:cahyo.untoro@if.its.ac.id), <sup>b,\*</sup> [leslie.anggraini@staff.its.ac.id](mailto:leslie.anggraini@staff.its.ac.id), <sup>c</sup> [maria.14117041@student.its.ac.id](mailto:maria.14117041@student.its.ac.id),

<sup>d</sup> [hesti.14117006@student.its.ac.id](mailto:hesti.14117006@student.its.ac.id), <sup>e</sup> [manas.14117028@student.its.ac.id](mailto:manas.14117028@student.its.ac.id)

\* Koresponden

#### INFO ARTIKEL

##### Sejarah artikel:

Menerima 16 Februari 2021  
Revisi 10 Maret 2021  
Diterima 27 Maret 2021  
Online 5 April 2021

##### Kata kunci:

corona  
COVID-19  
*k-means*  
pandemik  
pengelompokan

##### Keywords:

corona  
COVID-19  
*clustering*  
*k-means*  
pandemic

##### Style APA dalam menyitasi artikel ini:

Untoro, M. C., Anggraini, L., Andini, M., Retnosari, H., & Nasrulloh, M. A. (2021). Penerapan metode *k-means clustering* data COVID-19 di Provinsi Jakarta. *Teknologi: Jurnal Ilmiah Sistem Informasi*, 11(2), 59-68.

#### ABSTRAK

Wabah penyakit yang menyerang area pernafasan dan terdeteksi masuk di Indonesia pada awal tahun 2020 adalah Virus Corona (COVID-19). Penyebaran virus ini cukup mudah melalui *droplet* dari pasien terjangkit, sehingga penyebarannya sangat pesat. Penelitian ini dilakukan untuk mengklusterisasi data kasus COVID-19 di Provinsi Jakarta. Hal ini mengingat bahwa Jakarta merupakan titik awal kasus pertama Corona di Indonesia. Hingga kini, Jakarta menjadi salah satu penyumbang kasus COVID-19 terbesar di Indonesia dengan kasus positif per bulan Desember 2020 mencapai 154.000 jiwa dengan kesembuhan 139.000 jiwa. Pengelompokan dilakukan berdasarkan parameter pasien positif dan meninggal dari masing-masing kelurahan yang ada di Provinsi Jakarta. Penelitian ini menggunakan algoritma *k-means* untuk melakukan klusterisasi dalam penanganan kasus COVID-19 dengan 2 *cluster*. Persebaran data pada *cluster* 1 terdapat 173 data, dan 18 data pada *cluster* 2. Penggunaan *k-means* pada penelitian ini memberikan informasi daerah-daerah yang memiliki jumlah kasus positif dan tingkat kesembuhan tertinggi maupun terendah yang dapat dijadikan sebagai evaluasi dalam proses penanganan virus COVID-19.

#### ABSTRACT

The disease epidemic that attacked the respiratory area and was detected in Indonesia starting in early 2020 is the Corona Virus (COVID-19). This virus's spread is relatively easy, namely through droplets from infected patients, so that the spread is very rapid. This research was conducted to cluster the data on Covid-19 cases in Jakarta Province considering that Jakarta is the starting point for the first case of Corona in Indonesia and until now has become one of the most significant contributors to COVID-19 issues in Indonesia, namely as of December 2020 positive cases of Covid-19 reached 154,000. Souls with the healing of 139.0000 souls. The grouping was carried out based on positive and dead patients from each urban village in Jakarta Province. This study uses the *k-means Method* to cluster in the handling of COVID-19 cases with 2 clusters. Data distribution in cluster 1 consists of 173 data and 18 data in cluster 2. The use of *k-means* in this study provides information on areas with the highest and lowest number of positive cases and the highest and lowest cure rates that can be used as an evaluation in handling the Covid-virus 19.

Teknologi: Jurnal Ilmiah Sistem Informasi dengan lisensi CC BY NC SA.

## 1. Pendahuluan

Saat ini berbagai negara di seluruh dunia sedang dilanda pandemi virus COVID-19. COVID-19 menginfeksi sistem pernapasan manusia yang disebabkan oleh SARS-CoV-2 (Noviyanto, 2020; Dwitri,

Tampubolon, Prayoga, Zer, & Hartama, 2020). Virus ini ditemukan di Provinsi Wuhan, China yang diperkirakan pada akhir Desember 2019 untuk pertama kalinya muncul (Sindi, Ningse, Sihombing, Zer, & Hartama, 2020). Tidak dapat kita pungkiri juga bahwa penyebaran COVID-19 terus bergerak signifikan dikarenakan penyebaran virus ini cukup mudah, di mana penularannya melalui *droplet* dari pasien terjangkit. Penyebaran virus yang tergolong mudah ini menimbulkan ketakutan serta keresahan bagi masyarakat Indonesia (Rembulan, Wijaya, Palullungan, Alfina, & Qurthuby, 2020), sehingga pemerintah memberlakukan kebijakan-kebijakan salah satunya adalah Pembatasan Sosial Berskala Besar (PSBB) (Lesmana, Akbari, Rahman, & Gustian, 2020).

Sejak mulai terdeteksi pada awal tahun 2020, kasus COVID-19 di Indonesia bertambah begitu pesat. Provinsi Jakarta merupakan titik awal ditemukannya kasus Corona pertama di Indonesia. Provinsi Jakarta menjadi salah satu penyumbang kasus COVID-19 terbesar di Indonesia. Data pemantauan COVID-19 sejak tanggal 21 Januari 2020 hingga 4 April 2021 kasus positif COVID-19 mencapai 386.466 jiwa dengan kesembuhan 373.565 jiwa atau setara 96,7% (Jakarta, 2021).

Penelitian ini bertujuan untuk melakukan *clustering* terhadap data persebaran virus corona terutama di wilayah Provinsi Jakarta dengan menggunakan data-data yang telah disajikan pada portal <http://data.jakarta.go.id>. Data tersebut dapat dianalisis terkait peta penyebaran virus corona yang sebenarnya dari waktu ke waktu di Provinsi Jakarta.

*Clustering* adalah metode pada *data mining* yang digunakan untuk melakukan pengelompokan data yang memiliki karakteristik mirip ataupun sama menjadi sebuah *cluster* (Bastian, Sujadi, & Febrianto, 2018). Sedangkan untuk data yang memiliki karakteristik berbeda (Untoro & Buliali, 2018) akan dikelompokkan menjadi *cluster* baru (Gustientiedina, Adiya, & Desnelita, 2019). Metode *clustering* berbeda dengan metode *classification* dalam penyelesaian *dataset* (Untoro, et al., 2020). Proses algoritma *clustering* adalah mencari data pada seluruh set dan menjadikannya subkelompok, yang mana kesamaan data yang berada dalam *cluster* akan dimaksimalkan, dan kesamaan data yang berada dalam luar *cluster* akan diminimalkan (Taslim & Fajrizal, 2016).

Salah satu metode *clustering* yang sangat terkenal adalah algoritma *k-means*, karena *k-means* memiliki algoritma yang sederhana dan efisien. Sehingga *k-means* mudah untuk dipelajari (Gustientiedina, Adiya, & Desnelita, 2019). *k-means* merupakan metode yang cukup tangguh untuk digunakan di berbagai jenis data (Solichin & Khairunnisa, 2020). Oleh karena itu pada penelitian ini *k-means* dipilih untuk proses klasterisasi pada data persebaran COVID-19 di wilayah Jakarta. Hasil dari *clustering* ini nantinya diharapkan dapat menunjukkan perbedaan persebaran kasus COVID-19 di setiap wilayah pada Provinsi Jakarta yang dibagi berdasarkan kelurahannya. Informasi ini dapat digunakan untuk menentukan strategi yang dapat dilakukan untuk pencegahan penyebaran COVID-19 oleh pemerintah.

## 2. State of the Art

Pada penelitian sebelumnya belum ada penelitian yang berfokus untuk menerapkan metode *k-means* dalam mengklasifikasi persebaran kasus COVID-19 di wilayah Jakarta berdasarkan tingkat kelurahan. Sindi, Ningse, Sihombing, Zer, & Hartama (2020) menerapkan algoritma *k-Medoids* untuk mengelompokkan penyebaran COVID-19 di Indonesia. Hasil yang didapatkan adalah masyarakat cenderung terinfeksi COVID-19 di wilayahnya masing-masing. Penelitian tersebut menghasilkan jumlah *cluster* terbaik sebanyak 3 *cluster*. *Cluster* 1 memiliki 1 anggota, *cluster* 2 memiliki 2 anggota, dan *cluster* 3 memiliki 31 anggota.

Dwitri, Tampubolon, Prayoga, Zer, & Hartama (2020) menggunakan *k-means* untuk mengelompokkan data COVID-19 agar mengetahui persebaran penyakit di Indonesia. Persebaran *centroid* dari hasil *clustering* dengan *k-means* meliputi daerah Jakarta, Jawa Barat, dan Banten. Jakarta salah satu daerah yang memiliki persebaran data COVID-19 cukup tinggi, Jawa Barat potensial sedangkan Banten cukup potensial.

Suwandi, Situmorang, & Rakhmawati (2021) dalam penelitiannya mengungkapkan bahwa aplikasi *tracking* COVID-19 akan sangat membantu pemerintah dalam menanggulangi persebaran COVID-19 dengan tetap menjaga kerahasiaan data pribadi pengguna aplikasi. Hal tersebut sejalan dengan pengolahan dan penyimpanan data menurut SPBE berdasarkan PERPRES RI No. 95 Th 2018

bahwa segala data yang bersifat pribadi sangat dijaga sehingga aplikasi yang digunakan lebih aman dan nyaman bagi pengguna.

Fitriana, Saragih, & Hasyati (2018) dalam penelitiannya mengusulkan rancangan model sistem inteligensia bisnis untuk proses analisis dan pengukuran pemasaran produk di pabrik roti yang menggunakan metode *k-means*. Hasil penelitiannya memperoleh 83% pada *cluster* 1 yang merupakan kategori sisa roti rendah dan 17% pada *cluster* 2 merupakan sisa roti tinggi.

Indraputra & Fitriana (2020) dalam penelitiannya menggunakan *k-means* dengan bantuan *software* Microsoft Excel, Weka, dan KNIME dalam mengelompokkan tingkat terjangkit dan kematian yang disebabkan oleh COVID-19. Hasil diperoleh yaitu pada *cluster* 2 memiliki jumlah terjangkit dan meninggal lebih tinggi dibandingkan dengan *cluster* 1. Sehingga daerah-daerah *cluster* tersebut perlu diprioritaskan penanganannya.

Zubair, et al (2020) menyatakan bahwa penerapan *machine learning* dan *data science* dalam bidang kesehatan, salah satunya adalah untuk menentukan kelompok kualitas perawatan kesehatan di negara-negara dengan menggunakan *dataset* COVID-19. Proses *clustering* menggunakan algoritma *k-means* sudah teruji lebih efisien dalam menentukan pusat awal *cluster*.

Maulida (2018) menerapkan *data mining* dengan menggunakan *k-means* untuk mengelompokkan jumlah kunjungan wisatawan asing yang berada di Provinsi DKI Jakarta. Jumlah *cluster* dibentuk berdasarkan jumlah kunjungan wisata. Terdapat 3 *cluster*. *Cluster* pertama (C1) merupakan jumlah kunjungan wisatawan tinggi, *cluster* kedua (C2) adalah jumlah kunjungan wisatawan sedang, dan *cluster* ketiga (C3) adalah jumlah kunjungan wisatawan rendah. C1 berjumlah 1 data, C2 berjumlah 2 data dan C3 memiliki jumlah tertinggi yaitu 5 data. Hasil *clustering* menjadi catatan bagi pemerintah Provinsi DKI Jakarta untuk lebih memperhatikan tempat wisata yang berada di *cluster* C3 supaya dapat lebih maju seperti tempat wisata yang berada pada *cluster* C2 dan C1.

Budiman & Rudianto (2019) menggunakan metode *k-means* dalam mengelompokkan wilayah instansi pendaftar yang berada di daerah kota Serang. Hasil ini akan digunakan sebagai acuan daerah mana saja yang berpotensi mendatangkan mahasiswa baru, supaya jumlah mahasiswa baru setiap tahunnya dapat terus meningkat.

### 3. Metode Penelitian

#### 3.1. Pengumpulan data

Data yang digunakan dalam penelitian ini adalah data jumlah penduduk Jakarta dan data penyebaran COVID-19 di wilayah Jakarta. Tabel 1 diperoleh dari web Jakarta Open Data.

Tabel 1 *Dataset* COVID-19 Jakarta

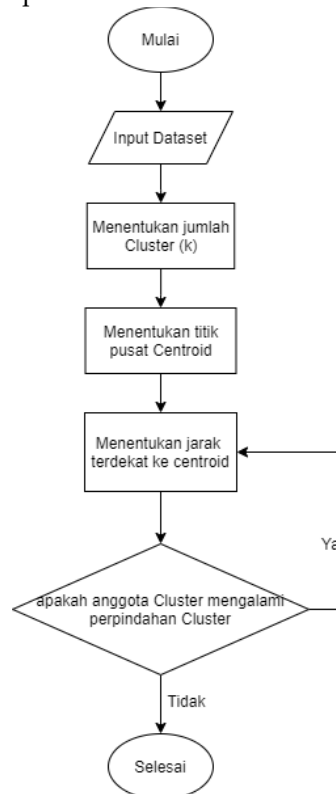
ID-KEL	KECAMATAN	KELURAHAN	SUSPEK	JUMLAH PENDUDUK	POSITIF	SEMBUH	MENINGGAL
3171011001	Pasar Minggu	Pejaten Timur	228	67.989	46	28	4
3171011002	Cilincing	Rorotan	165	50.167	10	4	1
3171011003	Grogol Petamburan	Tomang	302	8.843	91	59	2
.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....
3175101006	Cilandak	Gandaria Selatan	159	52.945	6	2	1
3175101007	Koja	Rawa Badak Selatan	215	53.360	37	26	2

#### 3.2. Clustering

Pada penelitian ini, algoritma *k-means* digunakan untuk *clustering*. Algoritma *k-means* merupakan algoritma yang bertujuan untuk mengelompokkan data menjadi beberapa *cluster* berdasarkan jarak terdekat (Indraputra & Fitriana, 2020). Gambar 1 merupakan diagram alir yang menggambarkan proses *clustering* yang dilakukan oleh *k-means*.

Jumlah *cluster* (*k*) atau nilai arbitrer ditentukan dengan menggunakan metode Elbow. Menurut Madhulatha (2012), metode Elbow merupakan metode yang digunakan untuk menentukan jumlah *cluster* terbaik dengan cara melihat titik yang membentuk siku, yang mana itu merupakan persentase dari hasil perbandingan antara jumlah *cluster*. Setelah nilai *k* didapatkan, maka akan diukur kualitasnya

menggunakan *Sillhouette Coefficient*. *Sillhouette Coefficient* akan mengukur seberapa dekat setiap objek dalam satu *cluster* dengan objek di *cluster* lain (Ogbuabor & Ugwoke, 2018). Menurut Kaufman & Rousseeuw (2005), *Sillhouette Coefficient* adalah kriteria untuk menilai atau mengukur baik tidaknya hasil *clustering* seperti yang disajikan pada Tabel 2.



Gambar 1. Proses *clustering* dengan *k-means*

Tabel 2. Kriteria penilaian *clustering* berdasarkan *Sillhouette Coefficient*

Nilai <i>Sillhouette Coefficient</i>	Penilaian
0,71 – 1,00	Struktur kuat
0,51 – 0,70	Struktur baik
0,26 – 0,50	Struktur lemah
≤ 0,25	Struktur buruk

#### 4. Hasil dan Pembahasan

Dalam melakukan penelitian ini penulis menggunakan bahasa pemrograman Python, Google Colab dan beberapa *library* untuk menjalankan algoritma *k-means*. *Dataset* yang digunakan adalah *dataset* jumlah penduduk Jakarta dan data persebaran COVID-19 di wilayah Jakarta. Proses yang pertama dilakukan yaitu melakukan proses *merge* pada *dataset1* dan *dataset2*. Proses *merge* ini dilakukan dengan cara menggabungkan *dataset1* dan *dataset2* yang memiliki nama\_kelurahan yang sama. Proses *merge* ini dapat dilihat pada Gambar 2.

```

In [3]: result = pd.merge(data1,
                        data2[['nama_kelurahan', 'positif', 'sembuh', 'meninggal']],
                        on='nama_kelurahan')

result.head()
  
```

Gambar 1. Proses *merge* pada *dataset1* dan *dataset2*

Selanjutnya akan dilakukan proses pembersihan data atau *data cleaning*. Proses *cleaning* akan dilakukan penghapusan subset data yang tidak menjadi perhatian pada penelitian atau menghapus data kosong yang berada di dalam *dataset*. Setelah proses *cleaning*, maka tersisa informasi kecamatan, nama\_kelurahan, jumlah\_penduduk, positif, sembuh, dan meninggal. Data yang dihasilkan disajikan pada Gambar 3.

Data yang akan diproses berjumlah 191 data seperti yang terlihat pada Gambar 3. Proses selanjutnya yaitu proses *slicing* pada *dataset* untuk membuat sebuah objek irisan. Kolom yang

dibutuhkan yaitu kolom 3 (positif) dan kolom 4 (sembuh). Perintah dan hasil dari proses *slicing* ini dapat dilihat pada Gambar 4.

	kecamatan	nama_kelurahan	jumlah-penduduk	positif	sembuh	meninggal
0	PENJARINGAN	PENJARINGAN	109567	156	77	5
1	PENJARINGAN	KAMAL MUARA	15051	16	12	1
2	PENJARINGAN	KAPUK MUARA	43025	18	14	2
3	PENJARINGAN	PEJAGALAN	90327	31	23	3
4	PENJARINGAN	PLUIT	55885	50	36	5
...	...	...	...	...	...	...
186	GROGOL PETAMBURAN	TANJUNG DUREN SELATAN	2761	40	26	5
187	GROGOL PETAMBURAN	TANJUNG DUREN UTARA	15047	18	9	1
188	KALIDERES	TEGAL ALUR	74349	23	17	1
189	GROGOL PETAMBURAN	TOMANG	8843	91	59	2
190	GROGOL PETAMBURAN	WIJAYA KUSUMA	90130	21	15	2

191 rows x 6 columns

Gambar 2. Dataset jumlah penduduk Jakarta dan data persebaran COVID-19 di Jakarta

In [8]:		Out[8]:	
<code>data = hasil.iloc[:, 3:5]</code>	<code>data</code>	positif	sembuh
		0	156
		1	16
		2	18
		3	31
		4	50
		...	...
		186	40
		187	18
		188	23
		189	91
		190	21
		191 rows x 2 columns	

(a)

(b)

Gambar 3. Proses untuk *slicing*: (a) Perintah *slicing*; dan (b) Hasil *slicing*

Langkah selanjutnya adalah mengubah variabel data *frame* menjadi *array* dan menstandarkan kembali ukuran variabel. Hal ini dilakukan supaya data yang telah dimodifikasi akan dapat kembali seperti jenis data awal sebelum data diubah menjadi *array*. Gambar 5 merupakan hasil dari menstandarkan ukuran variabel.

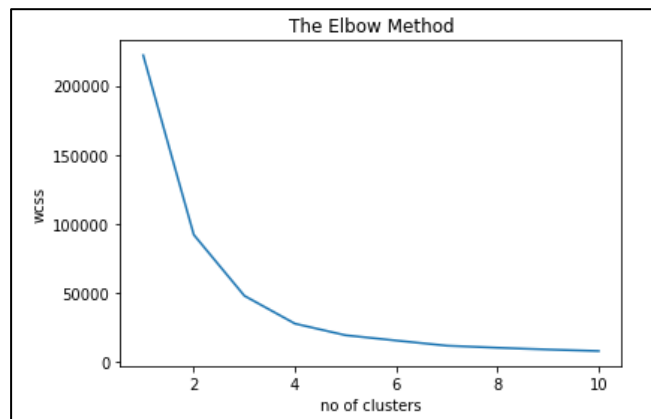
In [11]:	
<code>scaler = MinMaxScaler()</code>	<code>x_scaled = scaler.fit_transform(x_array)</code>
	<code>x_scaled</code>
Out[11]:	array([[0.88135593, 0.57462687], [0.09039548, 0.08955224], [0.10169492, 0.10447761], [0.17514124, 0.17164179], [0.28248588, 0.26865672], [0.29943503, 0.23134328], [0.64971751, 0.32089552], [0.06779661, 0.02238806], [0.07909605, 0.05223881], [0.23163842, 0.18656716], [0.97740113, 1. ], [0.20903955, 0.11940299], [0.08474576, 0.06716418], [0.24858757, 0.2238806 ], [0.36158192, 0.34328358], [0.11864407, 0.05223881], [0.12429379, 0.08955224], [0.20903955, 0.19402985], [0.23163842, 0.18656716],

Gambar 4. Hasil dari menstandarkan ukuran variabel

Penentuan nilai  $k$  ( $n\_clusters$ ) atau nilai arbitrer dilakukan dengan menggunakan metode Elbow. Setelah nilai  $k$  didapatkan, maka akan diukur kualitasnya menggunakan *Silhouette Coefficient*. *Silhouette*

*Coefficient* akan mengukur seberapa dekat setiap objek dalam satu *cluster* dengan objek di *cluster* lain (Ogbuabor & Ugwoke, 2018).

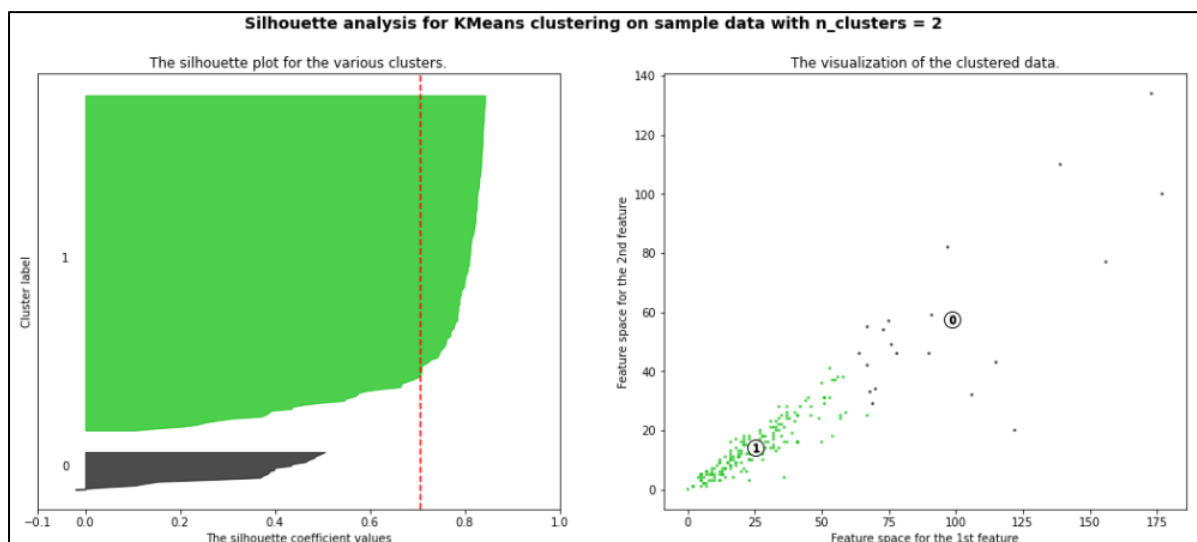
Dari Gambar 6 dapat dilihat bahwa nilai  $k$  yang efektif adalah  $k = 2$ , karena nilai  $k$  yang efektif menurut *Silhouette Coefficient* adalah saat skor *Silhouette* mendekati 1. *Silhouette Coefficient* dengan skor *Silhouette* berada diantara -1 hingga +1 (Ogbuabor & Ugwoke, 2018). Setelah itu, hal yang dilakukan adalah menentukan *cluster* dari hasil yang sudah didapatkan. Nilai *cluster* yang paling efektif yaitu  $k=2$  (Gambar 7 (b)). Semua hasil *cluster* disajikan pada Gambar 7 dan Gambar 8. Adapun perintah dan hasil dari menentukan dan mengonfigurasi fungsi *k-means* dapat dilihat pada Gambar 9.



Gambar 5 Penentuan nilai  $k$  dengan metode Elbow

```
For n_clusters = 2 The average silhouette_score is : 0.7063443964024457
For n_clusters = 3 The average silhouette_score is : 0.6243099117130568
For n_clusters = 4 The average silhouette_score is : 0.5292293264434884
For n_clusters = 5 The average silhouette_score is : 0.5000899474244881
For n_clusters = 6 The average silhouette_score is : 0.4736660947963217
For n_clusters = 7 The average silhouette_score is : 0.47951117319034303
For n_clusters = 8 The average silhouette_score is : 0.4333978810131924
For n_clusters = 9 The average silhouette_score is : 0.42402268018206
For n_clusters = 10 The average silhouette_score is : 0.4238148965577553
```

(a)

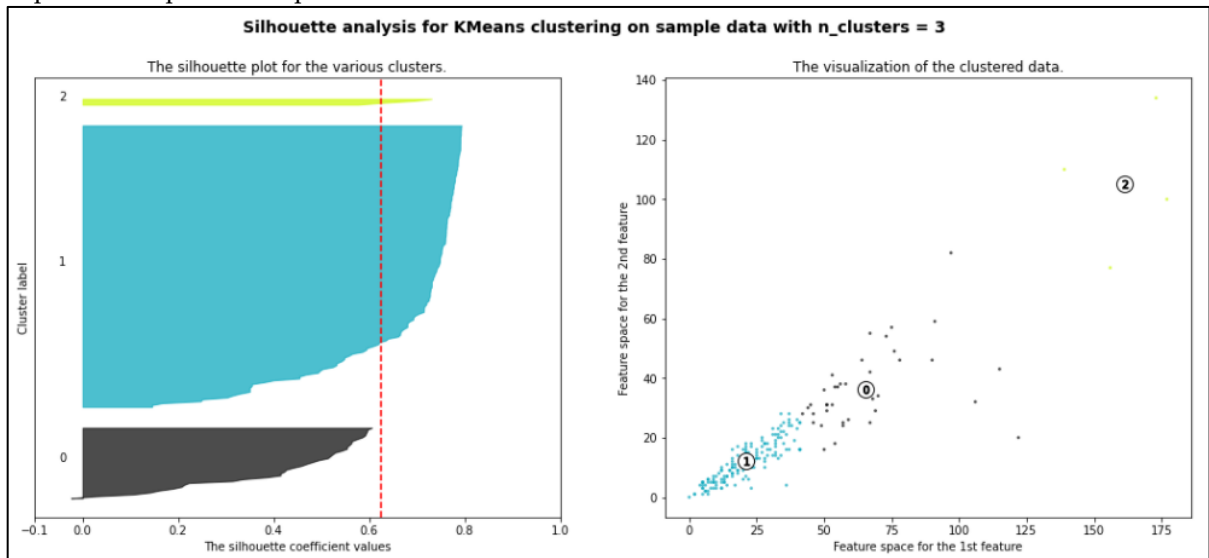


(b)

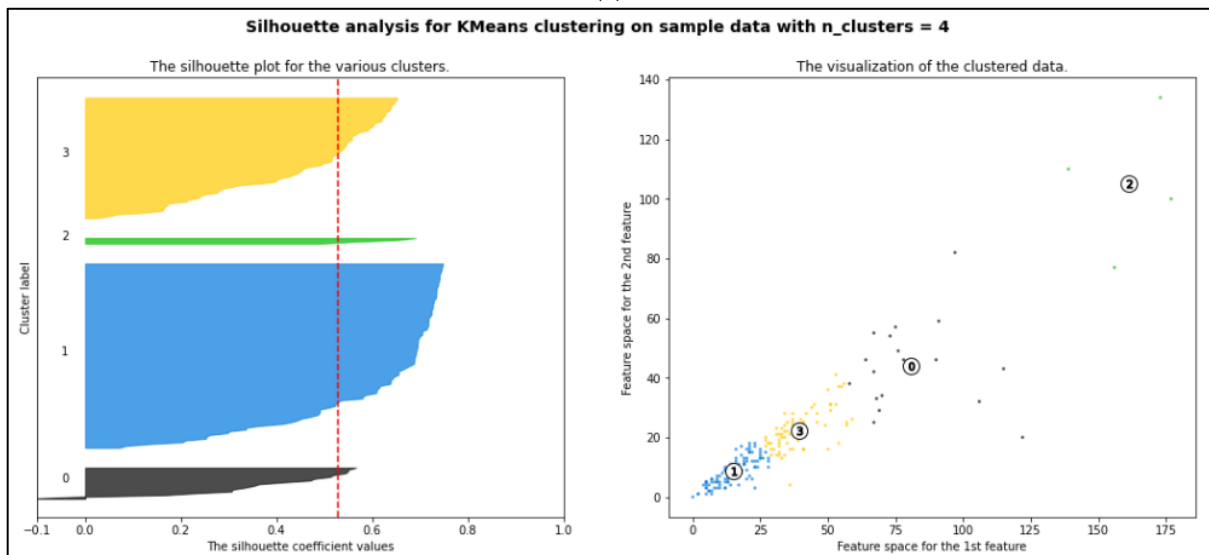
Gambar 6. Hasil *Silhouette Coefficient* (a) nilai *cluster* dari 2 – 10, (b) Jumlah *cluster* 2

Proses menampilkan pusat *cluster* dengan cara menampilkan *centroid* terlebih dahulu. Adapun cara untuk menampilkannya dapat dilihat pada Gambar 10. Dalam proses menampilkan data hasil *cluster* menambahkan kolom “klaster” ke dalam data *frame* pada *Frame Driver* dapat dilihat pada Gambar 11. Apabila sudah diklasterisasi, maka langkah selanjutnya yaitu menampilkan hasil *cluster* yang sudah didapatkan dari data hasil proses sebelumnya. Hasil *cluster* ini akan ditampilkan dalam

bentuk visualisasi yang dapat dilihat pada Gambar 12. Keseluruhan data divisualisasikan berdasarkan tiap *cluster* dapat dilihat pada Gambar 13.



(a)



(b)

Gambar 8. Hasil *Silhouette Coefficient* (a) Jumlah *cluster* 3, (b) Jumlah *cluster* 4.

```
In [7]: kmeans = KMeans(n_clusters = 2, random_state=10)
        kmeans.fit(x_scaled)
```

(a)

```
Out[7]: KMeans(n_clusters=2, random_state=10)
```

(b)

Gambar 9. Penentuan dan pengonfigurasi fungsi *k-means*: (a) Perintah; dan (b) Hasil (c)  $k=2$  (d)  $k=3$  (e)  $k=4$

```
In [13]: print(kmeans.cluster_centers_)
         [[0.1448679  0.1079717 ]
          [0.57627119 0.45024876]]
```

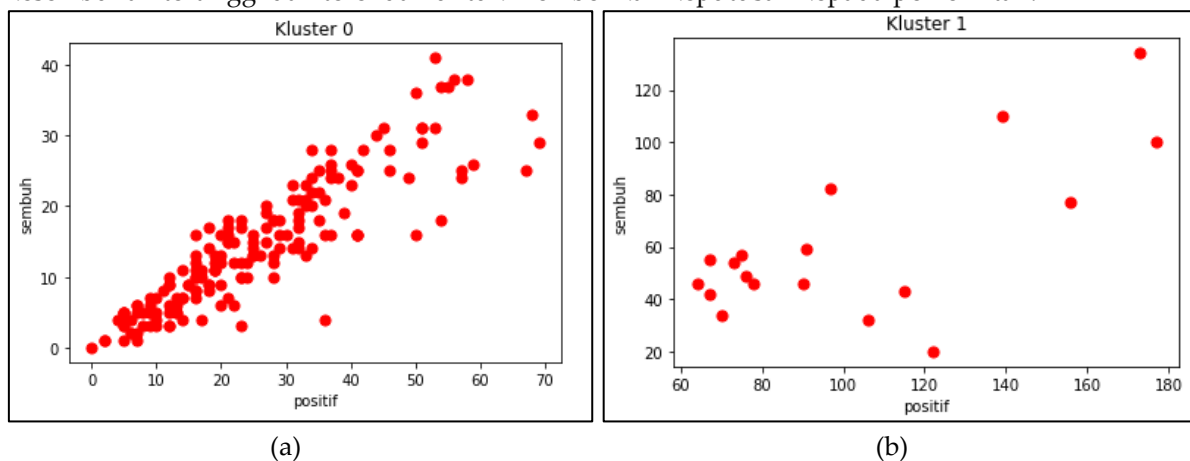
Gambar 10. Proses menampilkan pusat *cluster*





## 5. Kesimpulan

Kesimpulan yang berhasil didapatkan dari penelitian ini menggunakan data COVID-19 didapat 2 *cluster*. Pada *cluster* 0 terdapat 173 data, sedangkan pada *cluster* 1 terdapat 18 data. Hasil yang sudah didapatkan, dapat diketahui bahwa metode *k-means* dapat memberikan informasi persebaran daerah-daerah mana yang perlu diberikan penanganan COVID-19. Cempaka Putih Timur, Pegangsaan, Petamburan, Jembatan besi merupakan kelurahan yang perlu ditindaklanjuti dalam penanganan COVID-19. Persebaran data dapat digunakan sebagai informasi daerah-daerah yang memiliki jumlah kesembuhan tertinggi dan terendah untuk memberikan keputusan kepada pemerintah.



Gambar 13. Visualisasi data: (a) *cluster* 0; (b) *cluster* 1

## 6. Referensi

- Bastian, A., Sujadi, H., & Febrianto, G. (2018). Penerapan Algoritma k-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka). *Jurnal Sistem Informasi (Journal of Information System)*, 14(1), 26-32.
- Budiman, R., & Rudianto, R. (2019). Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means Clustering). *Protekinfo: Jurnal Pengembangan Riset dan Observasi Teknik Informatika*, 6(1), 6-14.
- Dwitri, N., Tampubolon, J. A., Prayoga, S., Zer, F. I., & Hartama, D. (2020). Penerapan Algoritma k-Means dalam Menentukan Tingkat Penyebaran Pandemi COVID-19 di Indonesia. *JurTI (Jurnal Teknologi Informasi)*, 4(1), 128-132.
- Fitriana, R., Saragih, J., & Hasyati, B. A. (2018). Perancangan Model Sistem Intelijensia Bisnis untuk Menganalisis Pemasaran Produk Roti di Pabrik Roti Menggunakan Metode Data Mining dan Cube. *Jurnal Teknologi Industri Pertanian*, 28(1), 113-126.
- Gustientiedina, G., Adiya, M., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru. *Jurnal Nasional Teknologi dan Sistem Informasi (TEKNOSI)*, 5(1), 17-24.
- Indraputra, R. A., & Fitriana, R. (2020). K-Means Clustering Data COVID-19. *Jurnal Teknik Industri*, 10(3), 275-282.
- Jakarta, P. (2021). *Data Pemantauan COVID-19*. Jakarta: Pemprov Jakarta. Retrieved April 4, 2021, from <https://corona.jakarta.go.id/id/data-pemantauan>
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.
- Lesmana, S., Akbari, A. F., Rahman, E. Y., & Gustian, D. (2020). Penerapan K-Means dalam Efektivitas Pembelajaran ELearning pada Masa Pandemi Covid-19. *Seminar Nasional Informatika 2020 (SEMNASIF 2020)*, (pp. 100-110).
- Madhulatha, T. S. (2012). An Overview on Clustering Methods. *IOSR Journal of Engineering*, 2(4), 719-725.
- Maulida, L. (2018). Penerapan Datamining dalam Mengelompokkan Kunjungan Wisatawan ke Objek Wisata Unggulan di Prov. DKI Jakarta dengan k-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 2(3), 167-174.
- Noviyanto, N. (2020). Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia. *Paradigma*, 22(2), 183-188.

- Ogbuabor, G., & Ugwoke, F. N. (2018). Clustering Algorithm for a Healthcare Dataset Using Silhouette Score Value. *International Journal of Computer Science & Information Technology (IJCSIT)*, 10(2), 27-37.
- Rembulan, G. D., Wijaya, T., Palullungan, D., Alfina, K. N., & Qurthuby, M. (2020). Kebijakan Pemerintah Mengenai Coronavirus Disease (COVID-19) di Setiap Provinsi di Indonesia Berdasarkan Analisis Klaster. *JIEMS (Journal of Industrial Engineering and Management Systems)*, 13(2), 74-86.
- Sindi, S., Ningse, W. R., Sihombing, I. A., Zer, F. I., & Hartama, D. (2020). Analisis Algoritma k-Medoids Clustering dalam Pengelompokan Penyebaran COVID-19 di Indonesia. *JurTI (Jurnal Teknologi Informasi)*, 4(1), 166-173.
- Solichin, A., & Khairunnisa, K. (2020). Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means. *Fountain of Informatics Journal*, 5(2), 52-59.
- Suwandi, S. I., Situmorang, A., & Rakhmawati, N. A. (2021). Analisis privasi data pengguna contact tracing application pengendalian COVID-19 di Indonesia berdasarkan PERPRES RI No. 95 tahun 2018 tentang sistem pemerintahan berbasis elektronik. *Teknologi: Jurnal Ilmiah Sistem Informasi*, 1(11), 46-58.
- Taslim, T., & Fajrizal, F. (2016). Penerapan algoritma k-mean untuk clustering data obat pada puskesmas rumbai. *Jurnal Teknologi Informasi & Komunikasi Digital Zone*, 7(2), 108-114.
- Untoro, M. C., & Buliali, J. L. (2018). Penanganan imbalance class data laboratorium kesehatan dengan Majority Weighted Minority Oversampling Technique. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 4(1), 23-29.
- Untoro, M. C., Praseptiawan, M., Widianingsih, M., Ashari, I. F., Afriansyah, A., & Oktafianto, O. (2020). Evaluation of Decision Tree, k-NN, Naive Bayes and SVM with MWMOTE on UCI Dataset. *Journal of Physics: Conference Series*, 1477.
- Zubair, M., Iqbal, M. A., Shil, A., Haque, E., Hoque, M. M., & Sarker, I. H. (2020). An Efficient K-means Clustering Algorithm for Analysing COVID-19.