

# Clustering Data Covid-19 Di Indonesia Menggunakan Intelligent K-Means

Veri<sup>1)</sup> Dyah Erny Herwindiati<sup>2)</sup> Teny Handhayani<sup>3)</sup>

<sup>1)2)3)</sup> Teknik Informatika, FTI, Universitas Tarumanaraga  
Jl. Letjen S Parman no. 1, Jakarta 11440 Indonesia

Email : [veri.535180004@stu.untar.ac.id](mailto:veri.535180004@stu.untar.ac.id)<sup>1)</sup>, [dyahh@fti.untar.ac.id](mailto:dyahh@fti.untar.ac.id)<sup>2)</sup>, [tenyh@fti.untar.ac.id](mailto:tenyh@fti.untar.ac.id)<sup>3)</sup>

## ABSTRACT

Indonesia pada terkena dampak wabah baru yakni virus covid-19. Covid-19 telah menjadi pandemi dikarenakan jumlah kasus di Indonesia yang terkonfirmasi terus meningkat. Di Jurnal ini menggunakan metode *Intelligent K-Means* untuk *clustering* data covid-19 yang ada di Indonesia. Setelah mendapatkan hasil kluster dari metode *Intelligent K-Means* hasil ini akan dibandingkan dengan hasil *clustering* yang dilakukan oleh pemerintah. Hasil evaluasi di jurnal ini menggunakan metode *Silhouette* yang di mana hasil ini evaluasi *silhouette* ini akan dibandingkan dengan metode *K-Means*.

## Kata kunci

*Intelligent K-Means, Covid-19, Silhouette, Clustering.*

## 1. Pendahuluan

Pada awal tahun 2020 ini, dunia dikejutkan dengan wabah virus *corona* (Covid-19) yang menginfeksi hampir seluruh negara di dunia. *World Health Organization* (WHO) semenjak Januari 2020 telah menyatakan dunia masuk ke dalam darurat global terkait virus ini. Ini merupakan fenomena luar biasa yang terjadi di bumi pada abad ke 21, yang skalanya mungkin dapat disamakan dengan Perang Dunia II, karena *event - event* skala besar (pertandingan-pertandingan olahraga internasional contohnya) hampir seluruhnya ditunda bahkan dibatalkan. Kondisi ini pernah terjadi hanya pada saat terjadi perang dunia saja, tidak pernah ada situasi lainnya yang dapat membatalkan acara-acara tersebut. WHO menyampaikan bahwa terhitung mulai tanggal 19 Maret 2020 sebanyak 214.894 orang terinfeksi virus *corona*, 8.732 orang meninggal dunia dan pasien yang telah sembuh sebanyak 83.313 orang.

Virus *Corona* atau *Covid- 19* ini masuk di Indonesia sejak akhir Februari 2020. Mengetahui adanya virus Covid-19 masuk di Indonesia pemerintah tidak tinggal diam. Banyak yang meragukan Indonesia

terkait penanganan virus *corona*, Presiden Indonesia buka suara. Jokowi meyakini sejak awal pemerintah telah benar-benar mempersiapkan segala sesuatunya. Termasuk rumah sakit 100 dengan ruangan standar isolasi. Jokowi juga mengatakan memiliki *reagen* yang cukup. “negara juga memiliki tim gabungan TNI atau Sipil dalam penanganan ini, (tim) ini yang belum pernah saya sampaikan sebelumnya”. Selain itu, Jokowi menyampaikan bahwa pemerintah juga memiliki standar operasional yang standarnya sama dengan standar internasional. Langkah-langkah telah dilakukan oleh pemerintah untuk dapat menyelesaikan kasus luar biasa ini, salah satunya adalah dengan mensosialisasikan gerakan *Social Distancing*. Konsep ini menjelaskan bahwa untuk dapat mengurangi bahkan memutus mata rantai infeksi Covid-19 seseorang harus menjaga jarak aman dengan manusia lainnya minimal 2 meter, dan tidak melakukan kontak langsung dengan orang lain, menghindari pertemuan massal. Kenyataannya banyak masyarakat yang menggampangkan atau bahkan menyepelekan hal ini , seperti contohnya masyarakat melihat ini sebagai hal yang enteng dan tidak mengindahkan himbauan pemerintah.

Penulis menggunakan Metode *Intelligent K-Means* pada jurnal ini untuk mendapatkan hasil *clustering* pada data covid-19 di Indonesia yang di mana hasil dari *clustering* ini akan dibandingkan dengan hasil *clustering* yang dilakukan oleh pemerintah.

## 2. Dasar Teori

### 2.1 *Intelligent K-Means*

*Intelligent K-Means* adalah varian dari *K-Means clustering* algoritma yang mengimplementasikan nilai rata-rata data untuk dihitung pusatnya. Algoritma *K-Means* menggunakan nilai artifisial yaitu sarana nilai fitur. *Intelligent K-Means (IK-Means)* Algoritma diusulkan oleh Boris Mirkin. *Intelligent K-Means* bukan para metrik *clustering* algoritma yang di rancang dari standar (para metrik) *K-Means*.

*Intelligent K-Means* jika dibandingkan dengan metode *K-Means* yakni *Intelligent K-Means* menentukan klusternya sendiri dari hasil rata-rata pada dataset dan pada metode *K-Means* manusia yang menentukan klusternya sendiri.

Ini adalah beberapa step untuk menggunakan *Intelligent K-Means* yakni sebagai berikut :

1. Hitung *Center of Mass (CoM)* atau *Grand Mean (GM)* menggunakan persamaan rata - rata.
2. Temukan Objek C1 yang memiliki jarak terjauh dari *CoM* menggunakan persamaan rata - rata.
3. Carilah objek lain C2 yang memiliki jarak terjauh dari C1 menggunakan persamaan rata - rata.
4. Hitung jarak objek lain terhadap *centroid* C1 dan C2 menggunakan persamaan rata - rata.
5. Kelompokkan objek, objek yang jaraknya paling dekat dengan C1 diberi label sebagai kluster S1, sedangkan objek lain dengan jarak dekat dengan C2 diberi label sebagai kluster S2.

$$\emptyset(S) = \frac{1}{l} \sum_{i=1}^l \emptyset(x_i) \quad (1)$$

$$d(i, j) = \frac{1}{\sqrt{(x_1 - x_2)^2 - (y_1 - y_2)^2}} \quad (2)$$

Keterangan :

$\emptyset(S)$  = *Center of Mass (CoM)*

$\emptyset(x_i)$  = *Kluster Centroid*

$l$  = ukuran *kernel* matriks

$x, y$  = vektor di dalam *kernel* matriks

$i, j$  = vektor indeks di dalam *kernel* matriks

Lima langkah pertama ini kemudian dimasukkan ke dalam langkah berikutnya yang adalah:

6. Untuk setiap objek dalam  $n$  kluster yang dihasilkan, temukan objek dengan jarak terjauh ke setiap kluster *centroid* sebagai kandidat *centroid* yang baru.
7. Temukan jarak rata-rata terbesar ke semua *centroid* dan gunakan objek itu sebagai *centroid* baru (C3).
8. Ulangi langkah 6 dan 7 hingga kluster konvergen (tidak ada objek yang mengubah kluster).

## 2.2 Silhouette

Tampilan grafis baru diusulkan untuk teknik partisi. Setiap kluster diwakili oleh apa yang disebut *Silhouette*, yang didasarkan pada perbandingan keketatan dan pemisahannya. *Silhouette* ini menunjukkan objek mana yang dia baik dalam kluster mereka, dan mana yang hanya di suatu tempat di antara kluster. Seluruh pengelompokan ditampilkan dengan menggabungkan *Silhouette* menjadi satu plot, memungkinkan apresiasi kualitas relatif dari kluster

dan gambaran umum konfigurasi data. Lebar *Silhouette* rata-rata memberikan evaluasi validitas pengelompokan, dan dapat digunakan untuk memilih jumlah kluster yang tepat.

Langkah dalam menghitung *Silhouette* di mulai dengan mencari jarak rata-rata data ke -i dengan semua data di klaster yang sama, di sini kita asumsikan data ke -i berada di klaster A. Rumus dari  $a(i)$  ditulis sebagai berikut :

$$a(i) := \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad (3)$$

Keterangan :

A = banyaknya data di klaster A

Selanjutnya menghitung nilai  $b(i)$  yang merupakan nilai minimum dari jarak rata-rata data ke-i dengan semua data di klaster berbeda. Sekarang, mari asumsikan klaster berbeda selain A dengan klaster C. Maka, perhitungan jarak rata-rata data ke -i dengan semua data di klaster C ditulis sebagai berikut:

$$d(i, C) := \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (4)$$

Keterangan :

C = banyaknya data di klaster C

Setelah menghitung  $d(i, C)$  untuk semua klaster  $C \neq A$ , selanjutnya memilih nilai jarak paling minimum sebagai nilai  $b(i)$ .

$$b(i) := \min_{C \neq A} d(i, j) \quad (5)$$

Jika klaster B memiliki nilai jarak minimum, maka  $d(i, B) = b(i)$  yang disebut sebagai tetangga dari data ke-i dan merupakan klaster terbaik kedua untuk data ke -i setelah klaster A. Setelah  $a(i)$  dan  $b(i)$  diketahui, maka proses terakhir menghitung *Silhouette*.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i) - b(i)\}} \quad (6)$$

Nilai  $s(i)$  berada antara -1 dan 1, di mana setiap nilai diinterpretasi sebagai berikut:

$s(i) \approx 1 \Rightarrow$  data ke -i digolongkan dengan baik (dalam A)

$s(i) \approx 0 \Rightarrow$  data ke -i berada di tengah antara dua klaster (A dan B)

$s(i) \approx -1 \Rightarrow$  data ke -i digolongkan dengan lemah (dekat ke klaster B daripada A)

Penafsiran nilai *Silhouette* bisa dilihat di bawah ini:

Tabel 1 Nilai *Silhouette*

<i>Silhouette</i>	Interpretasi
0.71 - 1.00	Struktur yang dihasilkan kuat
0.51 – 0.70	Struktur yang dihasilkan baik
0.26 – 0.50	Struktur yang dihasilkan lemah
≤ 0.25	Tidak terstruktur

### 3. Hasil Percobaan

Penulis menggunakan data set covid-19 yang sudah di *preprocessing* menggunakan metode *bloxplot* sebelum dimasukkan ke dalam metode *Intelligent K-Means* untuk di *clustering*. Hasil dari *clustering* data ini mendapatkan 6 kelompok kluster yang di mana terdiri dari kluster 0, kluster 1, kluster 2, kluster 3, kluster 4, dan kluster 5.

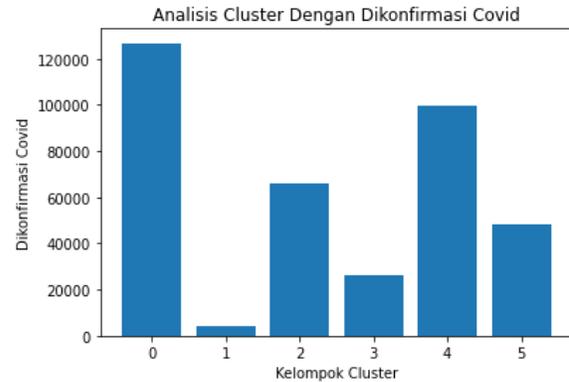
Setelah mendapatkan hasil kluster ini akan dievaluasi menggunakan metode *silhouette* yang di mana hasil evaluasi ini mendapatkan nilai 0.48 yang di mana nilai ini dianggap lemah pada *silhouette* dan pada *K-means* memiliki beberapa nilai yakni pada  $n\_clusters = 2$  memiliki nilai 0.743,  $n\_clusters = 3$  memiliki nilai 0.717,  $n\_clusters = 4$  memiliki nilai 0.673,  $n\_clusters = 5$  memiliki nilai 0.598, dan  $n\_clusters = 6$  memiliki nilai 0.602.

Pada perbandingan ini dapat dilihat *K-Means* memiliki nilai lebih tinggi dibandingkan dengan *Intelligent K-Means* tetapi semakin besar  $n\_clusters$  pada metode *K-Means* semakin kecil nilai *silhouette*, dibandingkan dengan *Intelligent K-Means* yang di mana hanya memiliki satu nilai *silhouette* saja.

Keuntungan menggunakan metode *Intelligent K-Means* yakni jika dibandingkan dengan metode *K-Means* yang di mana pada metode *K-Means* pada awal metode harus mengetahui berapa jumlah kluster yang ingin dibuat tetapi di metode *Intelligent K-Means* tidak perlu mengetahui berapa jumlah kluster tetapi sistemlah yang mencari otomatis jumlahnya.

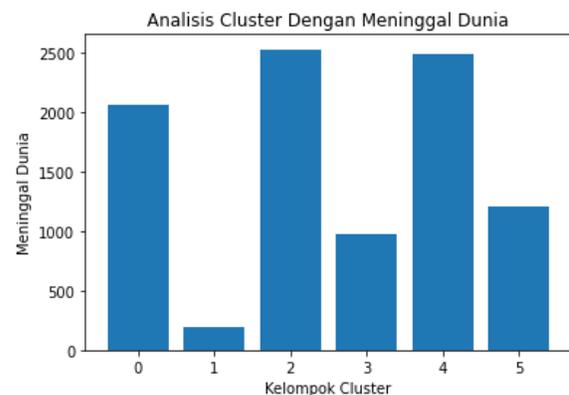
Pada hasil grafik analisis setiap kluster yang dibandingkan dengan rata-rata Dikonfirmasi Covid, Meninggal Dunia, Sembuh, Positif Covid-19, dan dibandingkan dengan Kepadatan Penduduk, dan Jumlah Penduduk per Puskesmas.

Pada analisis kluster dengan Dikonfirmasi Covid Kluster 0 memiliki rata-rata yang paling tinggi yakni 126778, pada Kluster 1 memiliki rata-rata 3995.266, pada Kluster 2 memiliki rata-rata 66252, pada Kluster 3 memiliki rata-rata 26017.15, pada Kluster 4 memiliki rata-rata 99913.4, pada Kluster 5 memiliki rata-rata 48278.78, yang di mana grafik ini bisa dilihat pada Gambar 1.



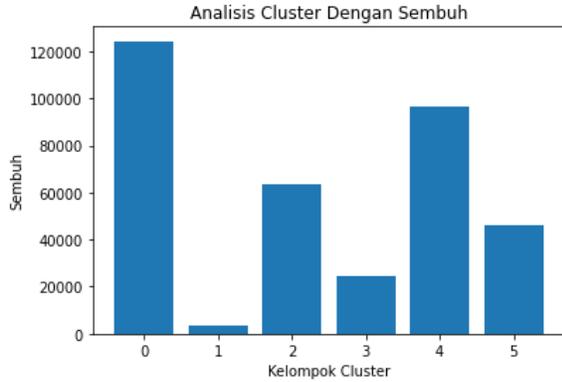
Gambar 1 Analisis Kluster Dengan Dikonfirmasi Covid

Pada analisis kluster dengan Meninggal Dunia Kluster 0 memiliki rata-rata 2067, pada Kluster 1 memiliki rata-rata 193.4738, pada kluster 2 memiliki rata-rata yang paling tinggi yakni 2533, pada Kluster 3 memiliki rata-rata 977.5, pada kluster 4 memiliki rata-rata 2492.6, pada kluster 5 memiliki rata-rata 1214, yang di mana grafik ini bisa dilihat pada Gambar 2.



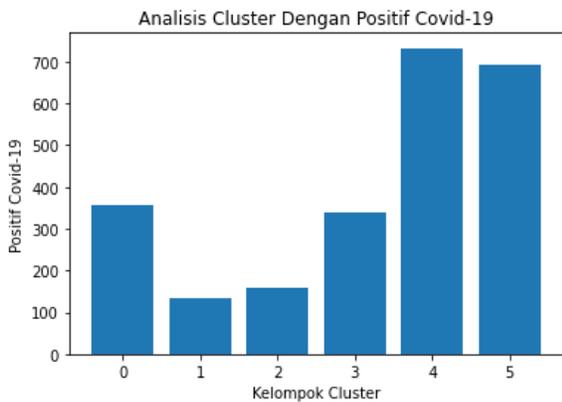
Gambar 2 Analisis Kluster Dengan Meninggal Dunia

Pada analisis kluster dengan Sembuh Covid-19 Kluster 0 memiliki nilai rata-rata yang paling tinggi yakni 124353, pada kluster 1 memiliki nilai rata-rata 3672.959, pada kluster 2 memiliki rata-rata 63561, pada kluster 3 memiliki rata-rata 24710.97, pada kluster 4 memiliki rata-rata 96688.6, pada kluster 5 memiliki rata-rata 46373.89, yang di mana grafik ini bisa dilihat pada Gambar 3.



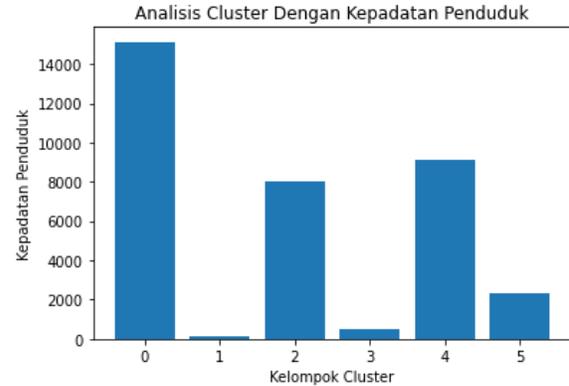
Gambar 3 Analisis Kluster Dengan Sembuh Covid

Pada analisis kluster dengan Positif Covid-19 Kluster 0 memiliki nilai rata-rata 358, pada kluster 1 memiliki nilai rata-rata 134.6856, pada kluster 2 memiliki rata-rata 158, pada kluster 3 memiliki rata-rata 339.3529, pada kluster 4 memiliki rata-rata yang paling tinggi yakni 732.2, pada kluster 5 memiliki rata-rata 690.8889, yang di mana grafik ini bisa dilihat pada Gambar 4.



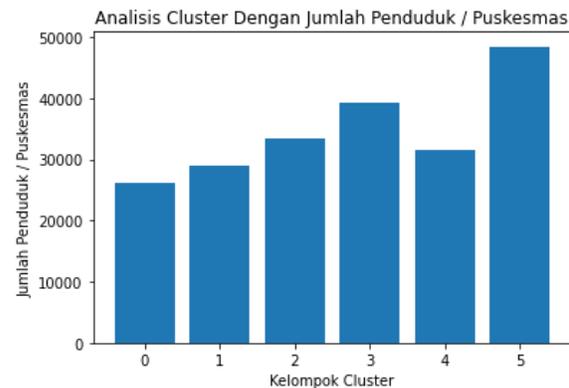
Gambar 4 Analisis Kluster Dengan Positif Covid-19

Pada analisis kluster dengan Kepadatan Penduduk Kluster 0 memiliki nilai rata-rata yang paling tinggi yakni 19045.4583, Kluster 1 memiliki nilai 109.6098, Kluster 2 memiliki nilai 400.8389, Kluster 3 memiliki nilai 323.9682, Kluster 4 memiliki nilai 5289.4005, Kluster 5 memiliki nilai 529.5238, yang di mana grafik ini bisa dilihat pada Gambar 5.



Gambar 5 Analisis Kluster Dengan Kepadatan Penduduk

Pada analisis kluster dengan Jumlah Penduduk per Puskesmas Kluster 0 memiliki nilai 26119.4857, Kluster 1 memiliki nilai 28960.4011, Kluster 2 memiliki nilai 33470.05, Kluster 3 memiliki nilai 39229.7329, Kluster 4 memiliki nilai 31543.6720, Kluster 5 memiliki nilai yang paling tinggi yakni 48448.4758, yang di mana grafik ini bisa dilihat pada Gambar 6.



Gambar 6 Analisis Kluster Dengan Jumlah Penduduk / Puskesmas

Hasil dari Peta Sebaran yang menggunakan Kelompok Kluster dapat dilihat pada Gambar 7. Pada Peta Sebaran ini dapat dilihat bahwa titik terparah untuk pulau yang terkena covid-19 yakni pulau Jawa. Pada pulau lainnya rata-rata memiliki nilai pada kelompok kluster yang sama yakni kluster satu tapi jika dibandingkan dengan pulau Jawa yang memiliki banyak kluster untuk covid-19 ini dapat disimpulkan bahwa pulau Jawa termasuk pulau yang sering menyebarkan covid-19 ini yang spesifiknya ada di provinsi Jakarta dan sekitarnya.



Gambar 7 Hasil Peta Sebaran Menggunakan Kelompok Kluster

Jika dibandingkan dengan hasil kluster pada Kawal Covid-19 ini dapat dilihat adanya persamaan yang di mana dipulau Jawa memiliki Index kewaspadaan yang sangat besar yang di mana bisa dilihat pada **Gambar 8**. Pada persamaan inilah dapat disimpulkan bahwa dipulau Jawa ini dampak dari covid-19 ini sangatlah tinggi dan harus diperhatikan oleh pemerintah.

Indeks Kewaspadaan - KawalCovid19 - Indonesia (Kab/Ko) - 18 September 2021



Gambar 8 Indeks Kewaspadaan dari Kawal Covid-19

#### 4. Kesimpulan

1. Metode *Intelligent K-Means* yang digunakan untuk pembuatan program ini dapat nilai evaluasi pada nilai *silhouette* yakni 0.498 yang di mana ini dianggap sebagai struktur yang lemah pada metode *silhouette*.
2. Hasil Peta Sebaran yang dibuat menggunakan *Geospasial* juga dapat dilihat adanya kesamaan jika dibandingkan dengan Indeks Kewaspadaan pada Kawal Covid-19 yang di mana dipulau Jawa yang memiliki kasus covid-19 terbanyak.
3. Hasil Perbandingan antara Metode *K-Means* dan Metode *Intelligent K-Means* pada nilai hasil *silhouette* ini dapat dilihat *K-Means* lebih unggul untuk nilai *silhouette* pada kluster berjumlah 6.

#### REFERENSI

- [1] Cahaya, Leonardo.; Hiryanto, Lely.; dan Handhayani, Teny. "Student graduation time prediction using intelligent K-Medoids Algorithm". 2017 3rd

International Conference on Science in Information Technology (ICSITech), 2017, pp. 263-266.

- [2] Handhayani, Teny.; dan Hiryanto, Lely. "Intelligent Kernel K-Means for Clustering Gene Expression". International Conference on Science and Computational Intelligence, Vol.59, Agustus, 2015.
- [3] Hidayati, Rahmatina.; Zubair, Anis.; Pratama, Aditya Hidayat.; dan Indana, Luthfi. "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering". Techno.COM. Vol. 20. No. 2. Mei, 2021.
- [4] Kalia, Baits Shalu Chandani. Analisis Penyebaran Berita Hoaks Pandemi Covid-19 di Bondowoso Melalui Facebook, <http://repository.unmuhjember.ac.id/id/eprint/8721>, 8 Maret 2022
- [5] Kawal Covid19. Tanya Jawab Seputar COVID-19. <https://kawalcovid19.id/faq>, 8 Maret 2022.
- [6] WHO. Coronavirus disease (COVID-19). <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19>, 8 Maret 2022.

Veri, Saat ini sebagai Mahasiswa Fakultas Teknologi Informatika jurusan Teknologi Informasi Universitas Tarumanagara.