

## Klasifikasi Data Covid 19 Di Wilayah Provinsi Indonesia Menggunakan Algoritma K-Means

*Classification of Covid-19 data in the provinces of Indonesia using the K-means algorithm*

**Muhamad Irfan Maulana - 312010045**

Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa  
Muhamadirfan10052000@gmail.com

### **Abstract**

*Covid 19 is a rapid spread virus that has a negative impact on the province of XYZ. With the Indonesian government policy, four zone policies such as the red, green, yellow, orange zone, which each zone has a different meaning. The algorithm used in this study is K-Means and K-Medoids. Algorithm-Means group data by dividing them into several clusters based on the same characteristics. Whereas the K-Medoids algorithm selects real objects to represent the cluster. In this study, the two algorithms were compared using one dataset. The comparison is done by looking at the value of Davies-Bouldin Index (DBI) on the rapidminer. The best result of K-Means DBI is 0.078 while K-Medoids gives a value of 0.250 which is divided into 2 clusters.*

**Keywords:** *Classification of Covid-19 data in the provinces of Indonesia using the K-means algorithm.*

### **Abstrak**

*Covid 19 merupakan virus yang menyebar cepat dan berdampak negatif bagi provinsi XYZ. Dengan kebijakan pemerintah Indonesia, kebijakan empat zona seperti zona merah, hijau, kuning, oranye, yang masing-masing zona memiliki arti yang berbeda. Algoritma yang digunakan dalam penelitian ini adalah K-Means dan K-Medoids. Algoritma-Berarti mengelompokkan data dengan membaginya menjadi beberapa kelompok berdasarkan karakteristik yang sama. Sedangkan algoritma K-Medoids memilih objek nyata untuk mewakili cluster. Dalam penelitian ini, kedua algoritma tersebut dibandingkan menggunakan satu dataset. Perbandingan dilakukan dengan melihat nilai Davies-Bouldin Index (DBI) pada rapidminer. Hasil terbaik dari K-Means DBI adalah 0,078 sedangkan K-Medoids memberikan nilai 0,250 yang dibagi menjadi 2 cluster.*

**Kata kunci:** Klasifikasi data covid 19 di wilayah provinsi indonesia menggunakan algoritma k-means.

### **Pendahuluan**

Pada 31 Desember 2019, WHO China Country Office melaporkan kasus penyebaran virus dikota Wuhan, Provinsi, China. Dan pada itu WHO memberikan nama virus tersebut Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-Co V-2) dan resmi menetapkan [enyakit tersebut dengan nama Novel Coronavirus pada manusia dengan sebutan Corona virus Diseases 2019 atau lebih dikenal sebutkan Covid 19. dimana jika tereserang penyakit ini akan menimbulkan gangguan pernapasan seperti flu dan gangguan berat lainnya[1].

Pada tanggal 2 Maret 2020, Indonesia untuk pertama kalinya mengonfirmasi kasus COVID-19. Pandemi COVID-19 yang terjadi ini membawa dampak yang serius pada tatanan kesehatan, perekonomian, dan sosial di indonesia[2].

Adapun algoritma yang digunakan untuk melakukan pengelompokan ialah K-Means. Algoritma K-Means adalah metode data mining yang termasuk kata lain tanpa supervisi ketika melakukan proses pemodelan, Metode K-Means pertama kali di perkenalkan oleh MacQueen JB pada tahun 1976. Metode ini adalah salah satu metode non hierarchi yang umum digunakan. Metode termasuk dalam teknik penyekatan (partition) yang membagi atau memisahkan objek k kedaerah bagian yang terpisah (Sibuea & Sapta, 2017). K-Means merupakan suatu algoritma pengklasteran yang cukup sederhana yang mempartisi database kedalam beberapa clasteran k. Algoritma cukup mudah untuk diimplementasikan dan dijalankan, relatif cepat, mudah disesuaikan dan banyak digunakan. Prinsip utama dari teknik ini adalah menyusun K buah partisi/pusat massa (centroid)/rata-rata (mean) dari sekumpulan data. Algoritma K-Means dimulai dengan pembentukan partisi kluster diawal kemudian secara iteraktif partisi claster ini diperbaiki hingga tidak terjadi perubahan yang signifikan pada partisi claster (Sibuea et al., 2017). Algoritma klasterisasi K-Means merupakan salah satu dari algoritma data mining yang populer di kalangan peneliti. Jika dibandingkan dengan algoritma klasterisasi lainnya, k-means juga masih cukup tangguh di berbagai jenis data (Solichin & Khairunnisa, 2020). [3].

#### Tinjauan Pustaka

Beberapa penelitian telah dilakukan untuk mengklasifikasi data covid 19 menggunakan algoritma k-means.

Penelitian Santoso dan Irsan (2020) Penelitian tersebut bertujuan mengelompokkan Provinsi yang rawan tindak kriminalitas agar dapat menentukan Provinsi mana saja yang membutuhkan ditingkatkan pengawasannya. Hasil penelitian yang dapat ditarik yaitu melalui perhitungan K-Means, didapatkan kelompok tindak kejahatan tinggi (cluster 1) terdapat 6 anggota, dan untuk kelompok tindak kejahatan rendah (cluster 2) ada 28 anggota [4].

Penelitian dengan kasus serupa juga dilakukan oleh Bu'ulolo dan Purba (2021) dengan algoritma yang berbeda untuk membentuk cluster zona penyebaran Covid-19 di Provinsi Sumatera Utara dengan algoritma K-Means menggunakan variabel positif, ODP, PDP dan meninggal. Hasil yang didapatkan dari penelitian ini yaitu diduplikatnya 3 cluster zona penyebaran [5].

Pada penelitian ini penulis ingin menerapkan data mining pada data kasus Covid-19 di Provinsi XYZ dengan menggunakan metode komprasi K-Means. data mining adalah suatu proses penggalian atau pengumpulan informasi penting dari suatu data yang besar [6].

Berdasarkan tinjauan pustaka tersebut, algoritma k-means terbukti dapat digunakan untuk mengklasifikasi data covid 19 di sertiap provinsi diwilayah indonesia.

#### Metode Penelitian

Dalam penelitian ini tahapan yang dilakukan adalah :

- 1) Tahap Pengumpulan Data,
- 2) Visualisasi,
- 3) Modelling, Pengujian Data dan Analisis Data dan
- 4) Kesimpulan

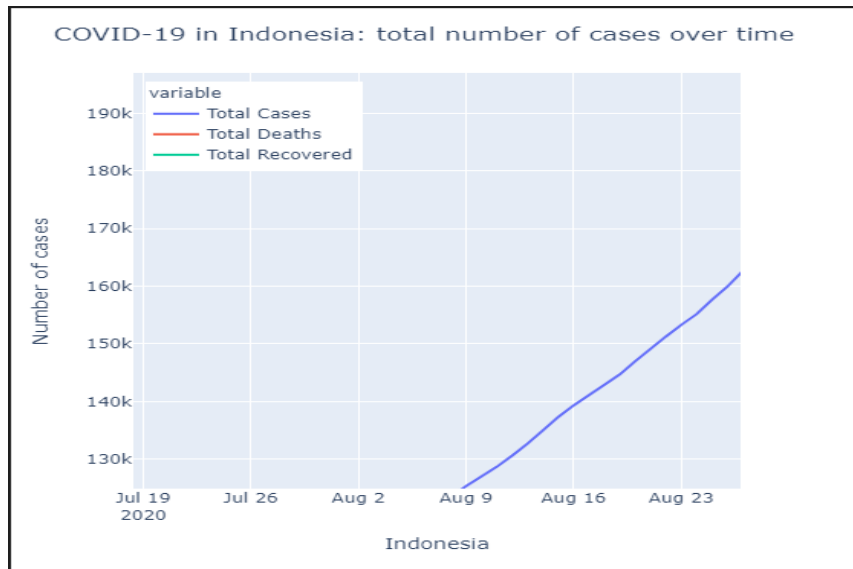
#### Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data mengenai kasus COVID-19 di Indonesia yang diambil dari website AtapData (atapdata.ai). Jumlah data yang digunakan mencapai 7098 data dari tanggal 1 Maret hingga 11 Oktober 2020 di 34 provinsi Indonesia. Data yang terdapat dalam dataset ini berupa : Date, Location ISO Code, Location, New Cases, New Deaths, New Recovered, New Active Cases, Total Cases, Total Deaths, Total Recovered, Total Active Cases, Location Level, City or Regency, Province, Country, Continent, Island, Time Zone, Special Status, Total Regencies, Total Cities, Total Districts, Total Urban Villages, Total Rural Villages, Area (km2), Population, Population

Density, Longitude, Latitude, New Cases per Million, Total Cases per Million, New Deaths per Million, Total Deaths per Million, Case Fatality Rate, Case Recovered Rate, Growth Factor of New Cases dan Growth Factor of New Deaths.

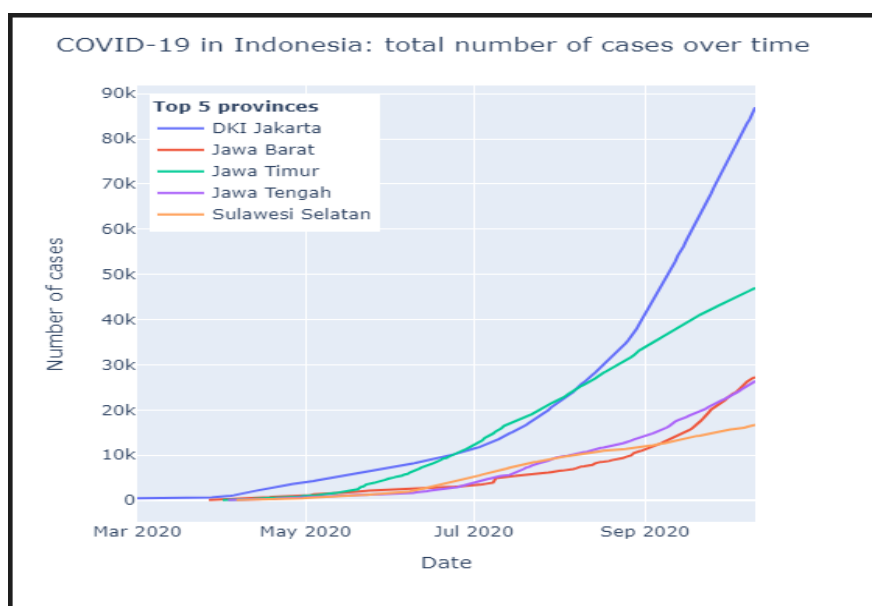
## Visualisasi

Visualisasi jumlah perkembangan kasus dengan line plot



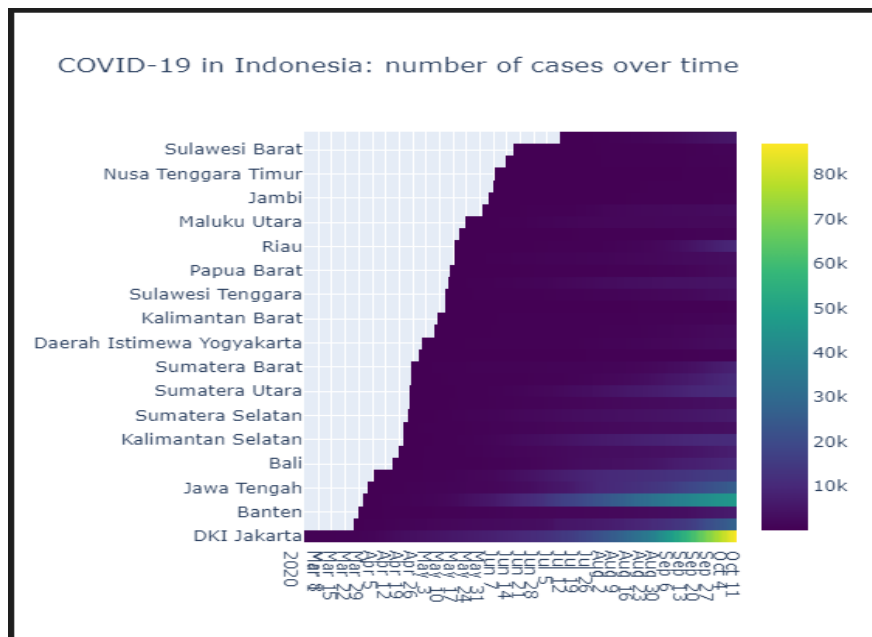
Pada visualisasi digambarkan kasus COVID-19 di Indonesia pada bulan Maret 2020 hingga Oktober 2020 mengalami peningkatan secara signifikan. Rata-rata kasus baru setiap harinya sebanyak 1.713 kasus. Jumlah kasus COVID-19 hingga 11 Oktober 2020 tercatat sebanyak 333.449 kasus dengan jumlah pasien sembuh sebanyak 255.027 orang dan jumlah pasien yang meninggal sebanyak 11.844 orang.

Visualisasi line plot 5 provinsi dengan jumlah kasus dan kematian tertinggi



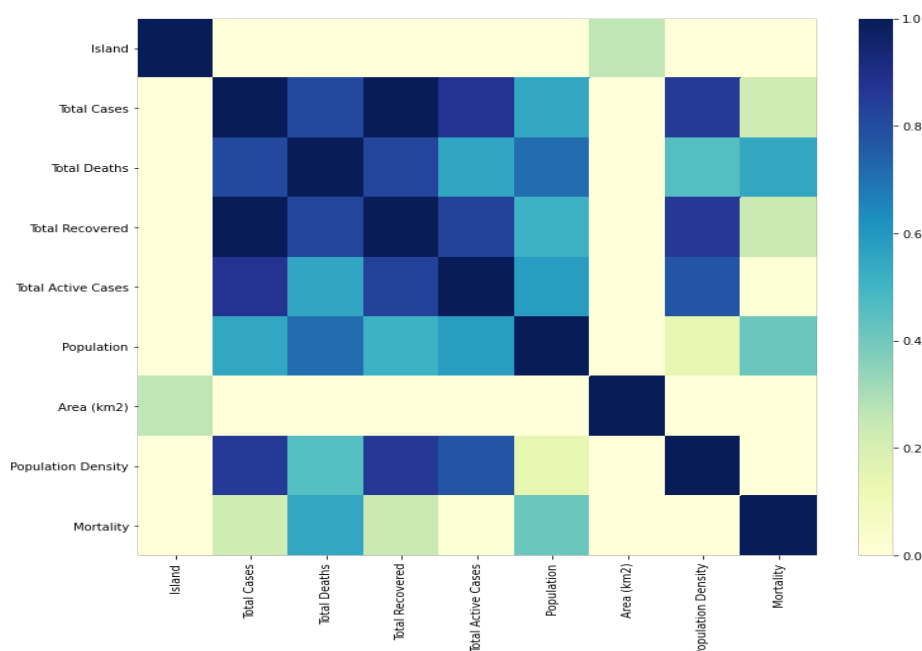
Pada visualisasi data 5 provinsi dengan jumlah kasus terkonfirmasi paling tinggi dari waktu ke waktu. Hingga 11 Oktober 2020 tercatat bahwa kasus terbanyak terjadi pada provinsi DKI Jakarta yakni sebanyak 86.963 kasus. Disusul provinsi Jawa Timur sebanyak 46.984 kasus, provinsi Jawa Barat sebanyak 27.236 kasus, provinsi Jawa Tengah sebanyak 26.401 kasus, dan di posisi kelima adalah provinsi Sulawesi Selatan sebanyak 16.738 kasus.

Visualisasi menggunakan heatmap untuk menunjukkan perkembangan dari penyebaran covid dalam beberapa minggu terakhir



Pada visualisasi diatas merupakan peta sebaran COVID-19 pada 34 provinsi di Indonesia menggunakan visualisasi heatmap yang menunjukkan tingkat keparahan atau tingginya kasus perhari nya dengan menunjukkan perubahan warna. Dalam visualisasi ini didapatkan DKI Jakarta memiliki perubahan warna yang signifikan yang menandakan DKI Jakarta mendapati tingkat kasus yang lebih banyak.

Menggunakan korelasi matriks



Dari hasil visualisasi korelasi matriks didapatkan korelasi antar fitur. Berdasarkan eksplorasi data yang telah diolah maka kami dapat melihat korelasi yang lebih kuat dengan jumlah kasus yaitu fitur berikut :

1. Total Cases
2. Total Recovered
3. Total Active Cases
4. Population Density
5. Total Deaths
6. Population
7. Mortality

**fitur-fitur ini yang akan digunakan untuk pengolahan data selanjutnya.**

Normalisasi, binning dan splitting dataset

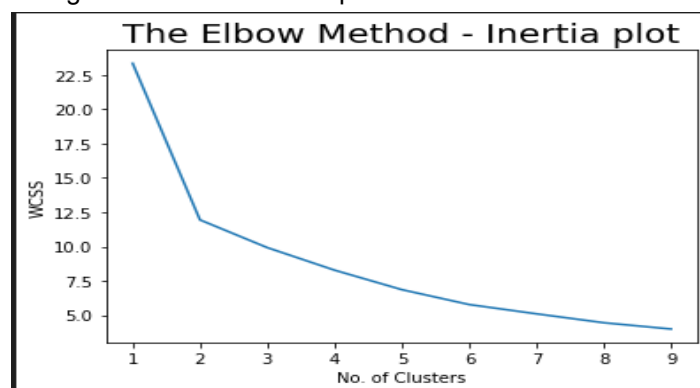
Splitting dataset dengan memilih fitur yang akan digunakan yaitu 'Total Cases', 'Total Recovered', 'Total Active Cases', 'Population Density', 'Total Deaths', 'Population', 'Mortality'

	Total Cases	Total Recovered	Total Active Cases	Population Density	Total Deaths	Population	Mortality
<b>Province</b>							
DKI Jakarta	86963	71454	13620	16334.31	1889	10846145	0.021722
Jawa Timur	46984	40490	3069	846.78	3425	40479023	0.072897
Jawa Barat	27236	17462	9244	1276.55	530	45161325	0.019460
Jawa Tengah	26401	18572	6322	1108.64	1507	36364072	0.057081
Sulawesi Selatan	16738	13891	2413	201.78	434	9426885	0.025929

Dari fitur yang dipilih akan dilakukan binning menjadi 7 bagian dan selanjutnya dilakukan normalisasi data sehingga mengurangi adanya outliers dan data lebih mudah dipahami dengan jarak antara 0-1

### Modelling, Pengujian Data dan Analisis Data

Menggunakan WCSS dengan Metode Elbow Graph untuk menentukan nilai cluster(k)

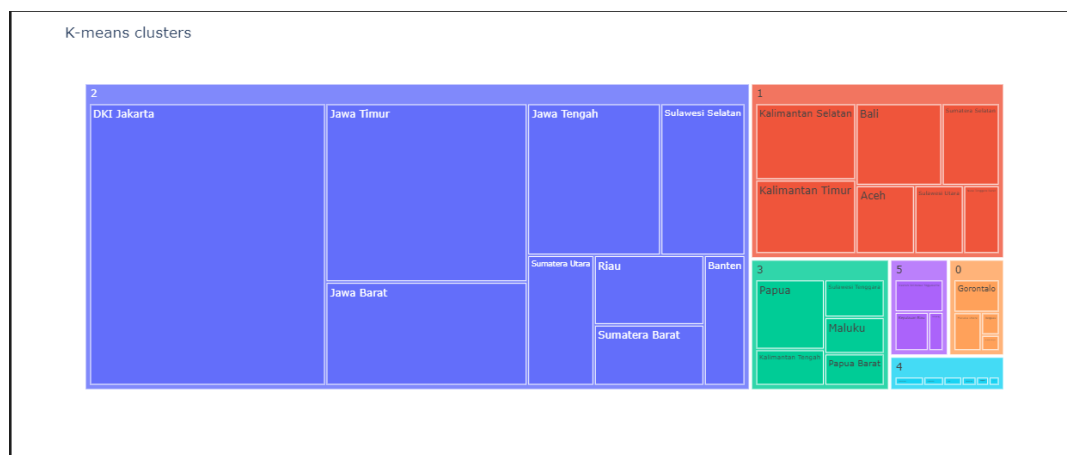


Pada visualisasi diatas dapat terlihat elbow pada 2 dan 6. dan dalam pengujian ini saya menggunakan k=6 agar lebih terlihat dan dapat lebih dapat insight yang diberikan

K-means Clustering dengan k=6

Province	Total Cases	Total Recovered	Total Active Cases	Population Density	Total Deaths	Population	Mortality	K-means
Lampung	0.166667	0.166667	0.166667	0.833333	0.166667	0.833333	0.666667	5
Daerah Istimewa Yogyakarta	0.333333	0.333333	0.500000	1.000000	0.500000	0.500000	0.500000	5
Kepulauan Riau	0.333333	0.333333	0.333333	0.833333	0.333333	0.166667	0.333333	5
Jambi	0.166667	0.000000	0.333333	0.333333	0.166667	0.333333	0.333333	4
Kepulauan Bangka Belitung	0.000000	0.000000	0.000000	0.333333	0.000000	0.000000	0.166667	4
Sulawesi Barat	0.166667	0.166667	0.166667	0.500000	0.000000	0.166667	0.000000	4
Kalimantan Barat	0.166667	0.166667	0.166667	0.166667	0.000000	0.666667	0.000000	4
Nusa Tenggara Timur	0.000000	0.000000	0.166667	0.666667	0.000000	0.666667	0.000000	4
Kalimantan Utara	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	4
Kalimantan Tengah	0.500000	0.500000	0.333333	0.000000	0.500000	0.166667	0.666667	3
Papua	0.666667	0.666667	0.833333	0.000000	0.500000	0.500000	0.166667	3
Sulawesi Tenggara	0.500000	0.333333	0.500000	0.166667	0.333333	0.333333	0.166667	3
Papua Barat	0.333333	0.166667	0.666667	0.000000	0.166667	0.000000	0.166667	3
Maluku	0.333333	0.333333	0.666667	0.166667	0.333333	0.166667	0.000000	3
Jawa Tengah	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	2
Jawa Timur	1.000000	1.000000	0.833333	0.833333	1.000000	1.000000	1.000000	2
Sumatera Utara	0.833333	0.833333	0.833333	0.666667	1.000000	1.000000	0.833333	2
Sulawesi Selatan	1.000000	1.000000	0.833333	0.666667	0.833333	0.833333	0.500000	2
Banten	0.666667	0.666667	0.666667	1.000000	0.666667	0.833333	0.500000	2
DKI Jakarta	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.333333	2
Riau	0.833333	0.833333	1.000000	0.333333	0.666667	0.833333	0.333333	2
Sumatera Barat	0.666667	0.666667	1.000000	0.666667	0.500000	0.666667	0.333333	2
Jawa Barat	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.166667	2
Sumatera Selatan	0.666667	0.666667	0.666667	0.333333	0.833333	0.833333	1.000000	1
Nusa Tenggara Barat	0.500000	0.500000	0.333333	0.833333	0.666667	0.666667	1.000000	1
Kalimantan Timur	0.833333	0.833333	0.833333	0.000000	0.833333	0.333333	0.833333	1
Kalimantan Selatan	0.833333	0.833333	0.500000	0.500000	0.833333	0.500000	0.833333	1
Sulawesi Utara	0.500000	0.666667	0.333333	0.666667	0.666667	0.333333	0.833333	1
Bali	0.833333	0.833333	0.500000	0.833333	0.833333	0.500000	0.666667	1
Aceh	0.666667	0.500000	0.666667	0.333333	0.666667	0.666667	0.666667	1
Bengkulu	0.000000	0.166667	0.000000	0.500000	0.166667	0.166667	1.000000	0
Sulawesi Tengah	0.000000	0.000000	0.000000	0.166667	0.166667	0.333333	0.833333	0
Maluku Utara	0.166667	0.333333	0.000000	0.166667	0.333333	0.000000	0.666667	0
Gorontalo	0.333333	0.500000	0.166667	0.500000	0.333333	0.000000	0.500000	0

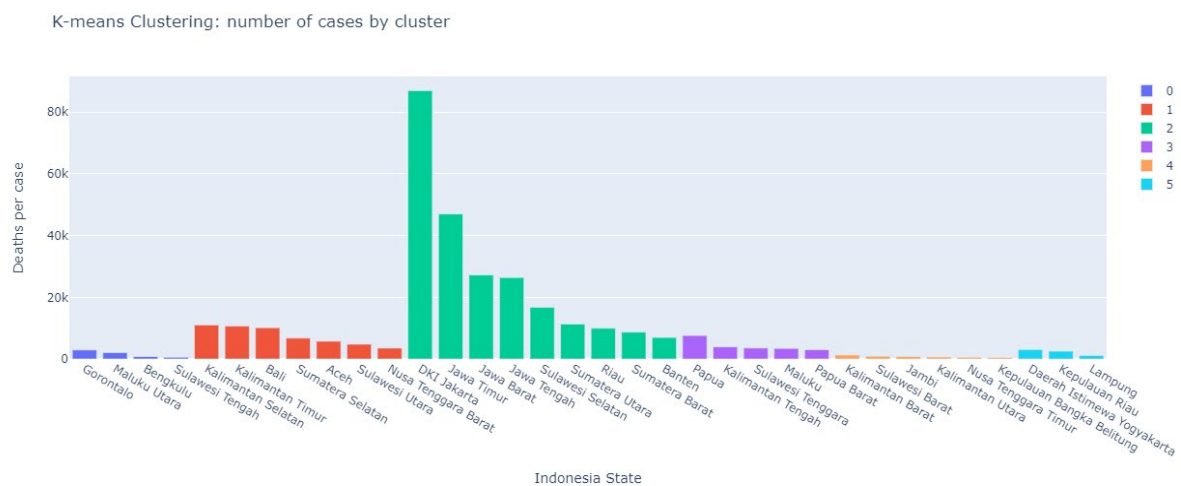
Visualisasi dengan Treemap hasil Klasterisasi Provinsi



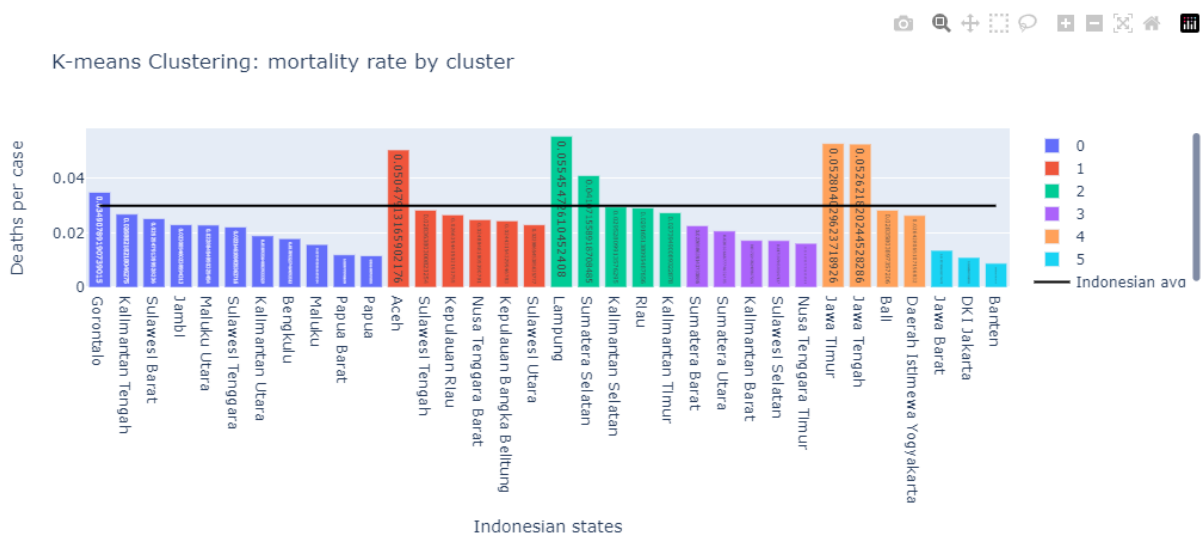
K-means clusters



Visualisasi Jumlah kasus aktif dari tiap provinsi dari masing-masing cluster menggunakan bar chart



Visualisasi Jumlah kemungkinan Kematian dari tiap provinsi dari masing-masing cluster.



Pada visualisasi diagram batang yang menunjukkan visualisasi tingkat kemungkinan kematian dari tiap-tiap provinsi menurut cluster yang telah dibagi sebelumnya. Dari ke-6 cluster diatas, cluster 1,2,4, dan 5 berada di atas rata-rata kemungkinan kematian di Indonesia dan Cluster 1 merupakan cluster yang memiliki tingkat kematian yang cukup besar dibandingkan cluster yang lain, Sedangkan provinsi yang memiliki tingkat kematian tergolong rendah terdapat pada cluster 2.

### **Analisis hasil pengujian**

Hasil Klasterisasi 6 klaster terbagi menjadi :

\* Klaster 1 (Cluster 0) : Bengkulu, Sulawesi Tengah, Kalimantan Tengah, Maluku Utara dan Gorontalo. Dengan rata-rata Tingkat kemungkinan Kematian 18% dan Total Kasus 10.348.

Klaster ini memiliki total kasus yang tergolong sedikit namun persentase tingkat kematiannya cukup banyak, dalam cluster ini perlu ditingkatkan terutama kinerja tenaga medis dan semangat pasien dalam menangani kasus COVID-19 serta tindakan-tindakan preventif yang dapat dilakukan dari masing-masing individu sesuai dengan protokol kesehatan yang ada.

\* Klaster 2 (Cluster 1) : Jawa Timur, Jawa Tengah, Sumatera Utara, Sulawesi Selatan, Riau, Banten, DKI Jakarta, Sumatera Barat dan Jawa Barat. Dengan rata-rata Tingkat kemungkinan Kematian 31% dan Total Kasus 241.327.

Klaster ini memiliki total kasus paling dan tingkat kematian paling tinggi maka perlu lebih ditingkatkan kembali kesadaran dari tiap individu-nya dalam menerapkan protokol Kesehatan dan melakukan berbagai tindakan preventif pencegahan COVID-19 lainnya, namun jika dilihat dari persentase kemungkinan kematian tidak berbanding jauh dengan Klaster 5 yang memiliki total kasus yang lebih sedikit, maka bisa disimpulkan bahwa tenaga medis dan pasien di kluster 2 sudah cukup baik dalam menangani kasus COVID-19.

\* Klaster 3 (Cluster 2) : Jambi, Kepulauan Bangka Belitung, Nusa Tenggara Timur, Kalimantan Utara, Kalimantan Barat, Sulawesi Barat. Dengan tingkat kemungkinan Kematian 7% dan Total Kasus 4687.

Klaster ini memiliki persentase tingkat kematian yang cukup rendah dan total kasus juga rendah maka Tindakan yang perlu dilakukan dalam cluster ini yaitu tindakan preventif dalam pencegahan COVID-19.

\* Klaster 4 (Cluster 3) : Sulawesi Tenggara, Papua, Papua Barat, Maluku. Dengan tingkat kemungkinan Kematian 5% dan Total Kasus 17.641.

Klaster ini memiliki persentase kematian yang cukup rendah dan tingkat kematian yang termasuk banyak, maka dapat disimpulkan dalam klaster ini tenaga medis dan pasien sudah cukup baik dalam menangani kasus COVID-19 namun kesadaran individu dalam melakukan Tindakan preventif masih rendah.

\* Klaster 5 (Cluster 4) : Nusa Tenggara Barat, Sumatera Selatan, Kalimantan Selatan, Sulawesi Utara, Aceh, Kalimantan Timur, Bali. Dengan tingkat kemungkinan Kematian 30% dan Total Kasus 52.704.

Klaster ini memiliki total kasus paling dan tingkat kematian ke-2 tertinggi maka perlu lebih ditingkatkan kembali kesadaran dari tiap individu-nya dalam menerapkan protokol Kesehatan dan melakukan berbagai tindakan preventif pencegahan COVID-19 lainnya, namun jika dilihat dari persentase kemungkinan kematian tidak berbanding jauh dengan Klaster 2 yang memiliki total kasus jauh lebih banyak bahkan 4x nya, namun persentase kemungkinan kematian nya tidak berbanding jauh maka bisa disimpulkan bahwa penanganan tenaga medis dan pasien COVID-19 perlu ditingkatkan kembali.



\* Klaster 6 (Cluster 5) : Lampung, Daerah Istimewa Yogyakarta, dan Kepulauan Riau. Dengan tingkat kemungkinan Kematian 8% Total Kasus 6742.

Klaster ini memiliki persentase tingkat kematian yang cukup tinggi jika dilihat dari total kasus tergolong rendah maka perlu ditingkatkan kinerja dalam penanganan COVID-19 baik dari tim tenaga medis maupun pasien COVID-19 nya.

## **Kesimpulan**

Pada karakteristik data diketahui bahwa rata-rata kasus baru setiap harinya sebanyak 1.713 kasus dengan kasus baru terbanyak yang terjadi pada 8 Oktober 2020 dengan jumlah kasus mencapai 4.850 kasus. Hingga 11 Oktober 2020 tercatat sebanyak 333.449 kasus dengan jumlah pasien sembuh sebanyak 255.027 orang dan jumlah pasien yang meninggal sebanyak 11.844 orang. DKI Jakarta menempati posisi pertama dengan jumlah kasus tertinggi yakni sebanyak 86.963 kasus. Provinsi dengan jumlah kasus paling banyak lainnya yaitu Jawa Timur, Jawa Barat, Jawa Tengah dan Sulawesi Selatan. Sedangkan provinsi dengan tingkat kematian paling tinggi akibat COVID-19 yaitu provinsi Jawa Timur. Dari hasil ini didapatkan bahwa Jumlah kematian pasien COVID-19 di suatu provinsi bukan acuan utama dalam menentukan klaster penanganan COVID-19, disamping menghitung jumlah kematian pasien ada indikator lain yang dapat mempengaruhi pembagian klasterisasi dalam penentuan penanganan COVID-19 di Indonesia Seperti tingkat kemungkinan kematian, jumlah kasus yang masih aktif, kepadatan penduduk dan luas area provinsi tersebut.

## Daftar Rujukan

- [1] Nur, A., Thohari, A., & Vernandez, A. B. (2020). Aplikasi Monitoring Kasus Coronavirus Berbasis Android. JTET (Jurnal Teknik Elektro Terapan) Polines, 9(1), 12–17.
- [2] Sibuea, F. L., & Sapta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. JURTEKSI (Jurnal Teknologi dan Sistem Informasi), 85 - 92.
- [3] Sindi, S., Ningse, W. R. O., Sihombing, I. A., Ilmi R.H.Zer, F., & Hartama, D. (2020). Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia. Jti (Jurnal Teknologi Informasi), 4(1), 166–173.
- [4] 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. <https://doi.org/10.21111/fij.v5i2.4905>
- [5] Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 epidemic. Tropical Medicine and International Health, 25(3), 278–280. <https://doi.org/10.1111/tmi.13383>
- [6] Dan, T. T., Sihwi, S. W., & Anggrainingsih, R. (2015). Implementasi Iterative Dichotomiser 3 Pada Data Kelulusan Mahasiswa S1 Di Universitas Sebelas Maret. JURNAL ITSMART, 84-91.
- [7] Hendini, A. (2016). Pemodelan UML Sistem Informasi Monitoring Penjualan Dan Stok Barang (Studi Kasus: Distro Zhezha Pontianak). JURNAL KHATULISTIWA INFORMATIKA, 107- 116.
- [8] Indraputra, R. A., & Fitriana, R. (2017). K-Means Clustering Data COVID-19. Jurnal Teknik Industri, 275-282.
- [9] Mardi, Y. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. Jurnal Edik Informatika, 213-219.