

# LAPORAN TUGAS AKHIR DATA MINING

## ANALISIS CLUSTERING KASUS COVID-19 DI INDONESIA

### MENGGUNAKAN ALGORITMA K-MEANS

Binti Fitrothul Khasanah<sup>1)</sup>, Dimas Wahyu Saputro<sup>2)</sup>, Viyonisa Syafa Sabila<sup>3)</sup>, Justin Tigor Hasonangan S<sup>4)</sup>, Marhanny Zahra N<sup>5)</sup>.

Program Studi Sains Data, Jurusan Sains, Institut Teknologi Sumatera

[binti.120450097@student.itera.ac.id](mailto:binti.120450097@student.itera.ac.id)<sup>1)</sup>, [dimas.120450081@student.itera.ac.id](mailto:dimas.120450081@student.itera.ac.id)<sup>2)</sup>,  
[viyonisa.120450027@student.itera.ac.id](mailto:viyonisa.120450027@student.itera.ac.id)<sup>3)</sup>, [justin.120450061@student.itera.ac.id](mailto:justin.120450061@student.itera.ac.id)<sup>4)</sup>,  
[marhanny.120450017@student.itera.ac.id](mailto:marhanny.120450017@student.itera.ac.id)<sup>5)</sup>

#### Abstrak

Adanya peningkatan kasus covid-19 maka perlu menganalisis kelompok untuk mengetahui daerah di Indonesia mana yang memiliki jumlah kasus covid-19 tertinggi. Dengan menerapkan metode data mining menggunakan algoritma *clustering K-Means*, penganalisaan kelompok dapat dilakukan. Data diambil dari *source* terbuka (Kaggle) dengan jumlah data 315.000 terdiri dari 9000 baris dan 35 kolom. Lalu data yang ada dikelompokkan berdasarkan daerahnya sehingga dapat diketahui daerah yang paling banyak jumlahnya, paling sedikit dan karakteristiknya. Hasil yang didapat memungkinkan pemerintah membuat sebuah kebijakan agar penyebaran kasus di daerah tersebut dapat diminimalisir. Ada enam (6) jenis *cluster* yang berhasil didapat dari data, masing-masing *cluster* menunjukkan tingkat resiko kematian dari berbagai daerah. Dari hasil tersebut disimpulkan bahwa jumlah kematian pasien covid-19 di provinsi tersebut bukan menjadi bahan perbandingan utama dalam menentukan *cluster* pengobatan covid-19. Selain menghitung jumlah kematian pasien, terdapat indikator lain yang dapat mempengaruhi distribusi kematian pasien. Pengelompokan untuk menentukan cara penanganan covid-19 di Indonesia, seperti: probabilitas kematian, jumlah kasus aktif, kepadatan penduduk dan luas provinsi.

**Kata Kunci :** Covid-19, data mining, *clustering K-Means*

## 1. Pendahuluan

### 1.1. Latar Belakang

Pada Akhir tahun 2019 tepatnya pada bulan desember, dunia dihebohkan dengan sebuah kejadian yang membuat banyak masyarakat resah yaitu dikenal dengan virus corona (covid-19). Kejadian tersebut bermula di Tiongkok, Wuhan. Peningkatan jumlah kasus COVID-19 terjadi begitu cepat dan sudah menyebar hampir ke seluruh dunia. Tanda dan gejala umum infeksi COVID-19 antara lain gejala gangguan pernapasan akut seperti demam, batuk dan sesak napas. Masa inkubasi rata-rata 5-6 hari dengan masa inkubasi terpanjang 14 hari. Pada kasus COVID-19 yang

berat dapat menyebabkan pneumonia, sindrom pernapasan akut, gagal ginjal, dan bahkan kematian. Tanda-tanda dan gejala klinis yang dilaporkan pada sebagian besar kasus adalah demam, dengan beberapa kasus mengalami kesulitan bernapas, dan hasil rontgen menunjukkan infiltrat pneumonia luas di kedua paru[1]. Sehingga untuk mencegah terjadinya penularan maka diberlakukan PSBB (Pembatasan Sosial Berskala Besar) di Indonesia, PSBB di bagi menjadi beberapa golongan tergantung jumlah kasus. Oleh karena itu masalah utamanya yaitu provinsi mana saja yang jumlah kasusnya tergolong banyak dan jumlah kasus yang tergolong sedikit. Sehingga dibutuhkan suatu program clustering untuk mendapatkan penggolongan provinsi berdasarkan banyaknya jumlah kasus yang terjadi dari provinsi tersebut.

### **1.2. Rumusan Masalah**

1. Bagaimana hasil dari pengklusteran ?
2. Provinsi mana saja yang termasuk dalam jumlah kasus yang tinggi?
3. Provinsi mana saja yang termasuk dalam jumlah kasus yang rendah?

### **1.3. Deskripsi Dataset**

Data yang digunakan dalam penelitian ini adalah data mengenai kasus COVID-19 di Indonesia yang diambil dari website AtapData ([atapdata.ai](https://atapdata.ai)) dan diambil dari Kaggle dengan jumlah data 315.000 terdiri dari 9000 baris dan 35 kolom yang diambil dari awal tahun 2020 sampai akhir tahun 2020. Data tersebut termasuk dengan data tabular dimana terdiri dari baris dan kolom. 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 (km<sup>2</sup>), 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.

## **2. Metode**

### **2.1. Gambaran Umum**

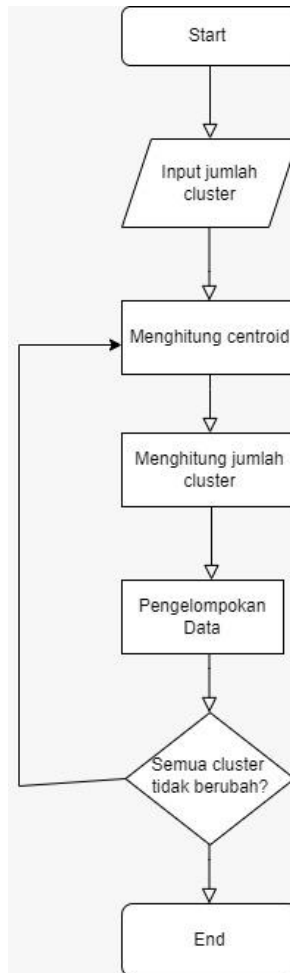
Pada paper ini menggunakan metode K-Means. K-Means merupakan salah satu algoritma dalam data mining yang digunakan untuk melakukan pengelompokan/clustering suatu data. Metode K-Means merupakan metode yang termasuk dalam algoritma clustering berbasis jarak yang membagi data ke dalam sejumlah cluster yang hanya bekerja pada atribut

numerik [2]. Metode K-Means merupakan salah satu metode *Unsupervised learning*. Tujuan dari algoritma K-Means adalah meminimalkan fungsi dengan meminimalkan variasi antar data dalam suatu cluster dan memaksimalkan variasi data pada cluster lainnya. Penelitian ini bertujuan untuk menerapkan algoritma K-Means untuk pengelompokan tingkat mortalitas covid 19 di Indonesia yang diperoleh dari website dataset Kaggle. Pengelompokan provinsi di Indonesia ke beberapa kluster tertentu sehingga dapat mengetahui daerah dengan jumlah kasus yang banyak dan yang sedikit. Hasil klasterisasi provinsi diharapkan dapat memberikan saran kepada pemerintah dalam menentukan aturan atau kebijakan terkait pembatasan kegiatan masyarakat atau kebijakan lainnya dalam mengatasi penyebaran COVID-19 dan memberi wawasan baru mengenai dampak Covid-19 pada masyarakat.

## **2.2. Flowchart Sistem**

Langkah-langkah algoritma K-Means yaitu[5]:

1. Memutuskan jumlah kluster k.
2. Menetapkan nilai-nilai awal k pusat cluster. Inisiasi dapat dilakukan dengan cara acak. Pusat-pusat cluster diberi nilai awal dengan angka-angka acak.
3. Menempatkan tiap-tiap item dataset yang jaraknya paling dekat dengan nilai centroid ke dalam centroid kluster tersebut.
4. Menghitung rata-rata nilai item dalam setiap kluster untuk dijadikan sebagai centroid yang baru.
5. Mengulang langkah 2 dan langkah 3 sehingga nilai centroid sama dengan nilai rata-rata item dalam kluster. Menggunakan euclidean distance untuk menghitung jarak antar titik. Alur dari algoritma K-Means dijelaskan pada Gambar 1.



Gambar 1. Alur Algoritma K-Means

### 2.3. Metrics Evaluation

Mengevaluasi dan validasi performa kluster dengan mengoptimalkan jumlah cluster dengan menggunakan teknik Davies Bouldin Index (DBI). Davies Bouldin Index adalah ukuran untuk mengevaluasi suatu pengelompokan. Davies Bouldin Index merupakan salah satu metode yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin. Davies Bouldin Index digunakan untuk mengevaluasi cluster secara umum berdasarkan kuantitas dan kedekatan antar anggota cluster. memiliki korelasi positif untuk kasus “dalam kelas” dan korelasi negatif untuk kasus “antar kelas”. Perhitungan nilai Davies Bouldin Index didasarkan pada rasio rasio cluster ke-i terhadap cluster ke-j. Semakin kecil nilai Davies Bouldin Index maka cluster yang dihasilkan semakin baik [3].

Metode Elbow merupakan salah satu metode untuk menentukan jumlah cluster yang tepat melalui persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik. Jika nilai cluster

pertama dengan nilai cluster kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka jumlah nilai cluster tersebut yang tepat [4].

#### 2.4. Cross-validation

Validasi silang dapat didefinisikan dalam pembelajaran yang diawasi. Dalam pembelajaran terawasi, selalu ada definisi galat yang jelas yang menghitung seberapa baik  $f(X)$  merupakan perkiraan dari  $y$ . Kita kemudian dapat menggunakan validasi silang untuk mendapatkan estimasi kesalahan model out-of-sample (atau generalisasi). Error ini bisa kita gunakan untuk melakukan pemilihan mode.

Dalam pembelajaran tanpa pengawasan, seperti pengelompokan, biasanya tidak ada definisi kesalahan yang jelas. Oleh karena itu, validasi silang juga tidak dapat digunakan untuk tujuan ini. Karena metode K-Means merupakan bagian dari *Unsupervised Learning* maka tidak melakukan Cross-validation.

### 3. Hasil dan Analisis

#### 3.1. Implementasi pada Bahasa Pemrograman Python

Dalam pengimplementasian pada bahasa Python, dapat dilihat sebagai berikut,

Processing Data

- a. *Import* berbagai library yang dibutuhkan, seperti *pandas* untuk melakukan manipulasi tabel, *matplotlib* untuk visualisasi data, serta *sklearn* untuk membuat model.

```
[ ] 1 import matplotlib.pyplot as plt
    2 import numpy as np
    3 import pandas as pd
    4 import plotly.express as px
    5 import plotly.graph_objects as go
    6 import seaborn as sns
    7 from sklearn.cluster import KMeans
    8 from sklearn.preprocessing import MinMaxScaler
```

Gambar 2. Library yang dibutuhkan

- b. *Import* dan *load* datasetnya

```
[ ] 1 data = pd.read_csv('/content/covid_19_indonesia_time_series_all.csv')
    2
    3 data.head()
```

Gambar 3 Load dataset

### c. Dataset Covid-19 di Indonesia

Date	Location ISO Code	Location	New Cases	New Deaths	New Recovered	New Active Cases	Total Cases	Total Deaths	Total Recovered	...	Longitude	Latitude	New Cases per Million	Total Cases per Million	New Deaths per Million	Total Deaths per Million	Case Fatality Rate	Case Recovered Rate
0 3/1/2020	ID-JK	DKI Jakarta	2	0	0	2	489	20	39	...	106.836118	-6.204699	0.18	45.09	0.0	1.84	4.09%	7.98%
1 3/2/2020	ID-JK	DKI Jakarta	2	0	0	2	491	20	39	...	106.836118	-6.204699	0.18	45.27	0.0	1.84	4.07%	7.94%
2 3/2/2020	IDN	Indonesia	2	0	0	2	2	0	0	...	113.921327	-0.789275	0.01	0.01	0.0	0.00	0.00%	0.00%
3 3/2/2020	ID-JS	Jawa Barat	3	0	0	3	12	5	120	...	107.603708	-6.920432	0.07	0.27	0.0	0.11	41.67%	1000.00%
4 3/2/2020	ID-RI	Riau	1	0	0	1	2	1	1	...	101.805109	0.511648	0.16	0.33	0.0	0.16	50.00%	50.00%

5 rows x 37 columns

Gambar 4 Dataset

### d. Pendefinisian Nama Kolom

```
[ ] 1 class ColumnData:
    2     date = 'Date'
    3     province = 'Province'
    4     island = 'Island'
    5     cases = 'Total Cases'
    6     deaths = 'Total Deaths'
    7     recovered = 'Total Recovered'
    8     actives_cases = 'Total Active Cases'
    9     population = 'Population'
    10    area = 'Area (km2)'
    11    mortality = 'Mortality'
    12    density = 'Population Density'
```

Gambar 5 Nama Kolom

### e. Memilih kolom yang akan digunakan

```
[ ] 1 data = data[[
    2     ColumnData.date,
    3     ColumnData.province,
    4     ColumnData.island,
    5     ColumnData.cases,
    6     ColumnData.deaths,
    7     ColumnData.recovered,
    8     ColumnData.actives_cases,
    9     ColumnData.population,
    10    ColumnData.area,
    11    ColumnData.density
    12    ]]
```

Gambar 6 kolom Terpilih

```
[35]: data
```

```
[35]:
```

	Date	Province	Island	Total Cases	Total Deaths	Total Recovered	Total Active Cases	Population	Area (km2)	Population Density
0	3/1/2020	DKI Jakarta	Jawa	489	20	39	430	10846145	664	16334.31
1	3/2/2020	DKI Jakarta	Jawa	491	20	39	432	10846145	664	16334.31
2	3/2/2020	NaN	NaN	2	0	0	2	265185520	1916907	138.34
3	3/2/2020	Jawa Barat	Jawa	12	5	120	-113	45161325	35378	1276.55
4	3/2/2020	Riau	Sumatera	2	1	1	0	6074100	87024	69.80
...	...	...	...	...	...	...	...	...	...	...
9954	12/31/2020	Sulawesi Tenggara	Sulawesi	7907	147	6696	1064	2635461	38068	69.23
9955	12/31/2020	Sulawesi Utara	Sulawesi	9671	310	7003	2358	2641884	13892	190.17
9956	12/31/2020	Sumatera Barat	Sumatera	23464	522	18030	4912	5519245	42013	131.37
9957	12/31/2020	Sumatera Selatan	Sumatera	11826	599	9364	1863	8217551	91592	89.72
9958	12/31/2020	Sumatera Utara	Sumatera	18149	679	15403	2067	14874889	72981	203.82

9959 rows x 10 columns

Gambar 7 Dataset sesudah pemilihan kolom

#### f. Memeriksa dan menghapus *Null Value*

```
[51]: data.isnull().sum()
```

```
[51]: Date                0
      Province           305
      Island             305
      Total Cases        0
      Total Deaths       0
      Total Recovered     0
      Total Active Cases  0
      Population          0
      Area (km2)          0
      Population Density  0
      dtype: int64
```

Menghapus jika ada data yang kosong

```
[9]: data = data.dropna(axis=0, how="any")
```

Gambar 8 Penghapusan Null Value

#### g. Melihat dataset yang sekarang

```
[59]: data
```

```
[59]:
```

	Date	Province	Island	Total Cases	Total Deaths	Total Recovered	Total Active Cases	Population	Area
0	3/1/2020	DKI Jakarta	Jawa	489	20	39	430	10846145	
1	3/2/2020	DKI Jakarta	Jawa	491	20	39	432	10846145	
3	3/2/2020	Jawa Barat	Jawa	12	5	120	-113	45161325	
4	3/2/2020	Riau	Sumatera	2	1	1	0	6074100	
5	3/3/2020	DKI Jakarta	Jawa	493	20	39	434	10846145	
...	...	...	...	...	...	...	...	...	...
9954	12/31/2020	Sulawesi Tenggara	Sulawesi	7907	147	6696	1064	2635461	
9955	12/31/2020	Sulawesi Utara	Sulawesi	9671	310	7003	2358	2641884	
9956	12/31/2020	Sumatera Barat	Sumatera	23464	522	18030	4912	5519245	
9957	12/31/2020	Sumatera Selatan	Sumatera	11826	599	9364	1863	8217551	
9958	12/31/2020	Sumatera Utara	Sumatera	18149	679	15403	2067	14874889	

9654 rows x 10 columns

Gambar 9 Dataset setelah penghapusan Null Value

#### h. Membuat format data dan membuat kolom baru

Membuat format data pada tanggal menjadi date.

```
[39]: data[ColumnData.date] = pd.to_datetime(data.Date, infer_datetime_format=True).dt.date
```

Membuat kolom baru yaitu 'Mortality'

```
[41]: data[ColumnData.mortality] = data[ColumnData.deaths] / data[ColumnData.cases]
```

Gambar 10 Pengkodean membuat format dan kolom

i. Dataset final yang akan digunakan

[42]: data

[42]:

	Date	Province	Island	Total Cases	Total Deaths	Total Recovered	Total Active Cases	Population	Area (km2)	Population Density	Mortality
0	2020-03-01	DKI Jakarta	Jawa	489	20	39	430	10846145	664	16334.31	0.040900
1	2020-03-02	DKI Jakarta	Jawa	491	20	39	432	10846145	664	16334.31	0.040733
2	2020-03-02	NaN	NaN	2	0	0	2	26518520	1916907	138.34	0.000000
3	2020-03-02	Jawa Barat	Jawa	12	5	120	-113	45161325	35378	1276.55	0.416667
4	2020-03-02	Riau	Sumatera	2	1	1	0	6074100	87024	69.80	0.500000
...	...	...	...	...	...	...	...	...	...	...	...
9954	2020-12-31	Sulawesi Tenggara	Sulawesi	7907	147	6696	1064	2635461	38068	69.23	0.018591
9955	2020-12-31	Sulawesi Utara	Sulawesi	9671	310	7003	2358	2641884	13892	190.17	0.032055
9956	2020-12-31	Sumatera Barat	Sumatera	23464	522	18030	4912	5519245	42013	131.37	0.022247
9957	2020-12-31	Sumatera Selatan	Sumatera	11826	599	9364	1863	8217551	91592	89.72	0.050651
9958	2020-12-31	Sumatera Utara	Sumatera	18149	679	15403	2067	14874889	72981	203.82	0.037413

9959 rows x 11 columns

Gambar 11 Dataset yang akan digunakan

j. Normalisasi, binning dan splitting dataset

	Total Cases	Total Recovered	Total Active Cases	Population Density	Total Deaths	Population	Mortality
Province							
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

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
Province							
DKI Jakarta	1.0	1.0	1.000000	1.000000	1.000000	1.000000	0.333333
Jawa Timur	1.0	1.0	0.833333	0.833333	1.000000	1.000000	1.000000
Jawa Barat	1.0	1.0	1.000000	1.000000	1.000000	1.000000	0.166667
Jawa Tengah	1.0	1.0	1.000000	1.000000	1.000000	1.000000	1.000000
Sulawesi Selatan	1.0	1.0	0.833333	0.666667	0.833333	0.833333	0.500000

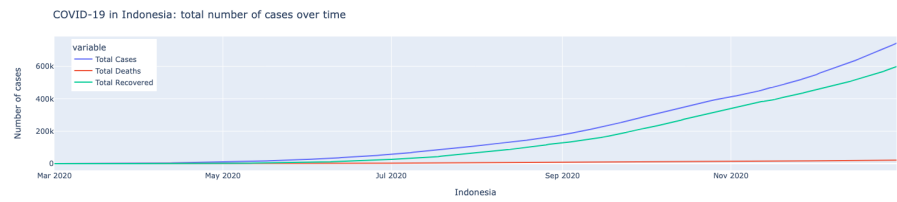
Dari fitur yang dipilih akan dilakukan binning menjadi 7 bagian dan selanjutnya dilakukan normalisasi data sehingga mengurangi adanya outliers dan data lebih mudah dipahami dengan jarak antara 0-1

## Visualisasi Data

k. Visualisasi jumlah perkembangan kasus dengan line plot

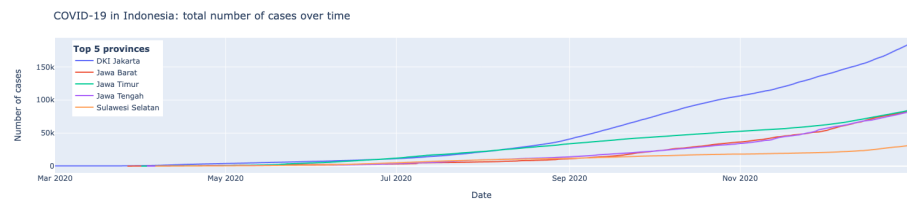
Pada visualisasi digambarkan kasus COVID-19 di Indonesia pada bulan Maret 2020 hingga Oktober 2020 mengalami peningkatan secara signifikan. Rata-rata kasus baru setiap harinya sebanyak 1.713 kasus. Jumlah kasus COVID-19 hingga 11 Oktober 2020 tercatat sebanyak 333.449 kasus dengan jumlah pasien sembuh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





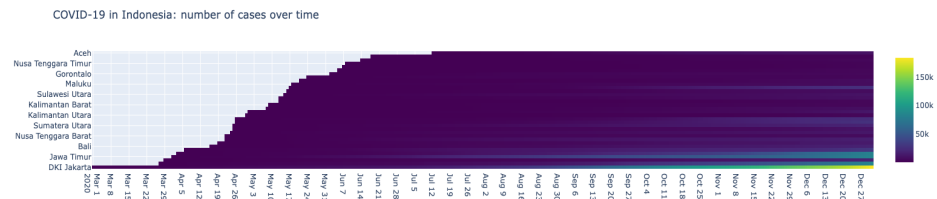
Gambar 12 Line plot Jumlah Perkembangan Kasus

Visualisasi line plot 5 provinsi dengan jumlah kasus dan kematian tertinggi



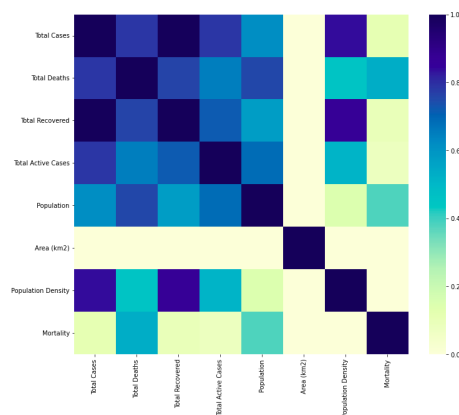
Gambar 13 Line Plot Lima Provinsi Tertinggi

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



Gambar 14 Visualisasi menggunakan heatmap untuk menunjukkan perkembangan dari penyebaran covid dalam beberapa minggu terakhir

Menggunakan korelasi matriks



Gambar 15 Korelasi Matriks

## Modelling, Pengujian Data dan Analisis data

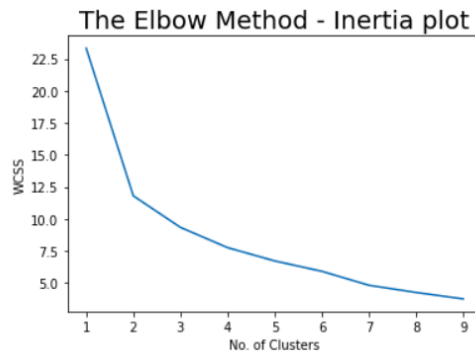
### 1. Load dataset yang akan digunakan

```
#selecting_features_
X = d[['Mortality', 'Total_Cases', 'Total_Active_Cases', 'Population_Density', 'Population', 'Total_Deaths']]
```

Gambar 16 Load dataset

### m. 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.



Gambar 17 Elbow Method

[20]:

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

Gambar 18 K Means Clustering dengan k=6



- 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.

#### 4. 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.

#### Referensi

- [1] Putri, R. N. (2020). Indonesia dalam Menghadapi Pandemi Covid-19. *Jurnal Ilmiah Universitas Batanghari Jambi*, 20(2), 705. <https://doi.org/10.33087/jiubj.v20i2.1010>
- [2] Witten, et al., 2012, *Data Mining Practical Machine Learning Tools and Technique*, 2nd Edition, Morgan Kaufmann, San Francisco.
- [3] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM J.*, vol. 10, no. 3, pp. 1099–1103, 2021.
- [4] N. Putu, E. Merliana, and A. J. Santoso, "Analisa penentuan jumlah cluster terbaik pada metode KMeans," *Seminar Nasional Multi Disiplin Ilmu*, pp. 978–979.
- [5] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021.

#### Lampiran

Kode Pemrograman, terlampir pada .zip