

## Klasterisasi Pasien COVID-19 Berdasarkan Komorbiditas Menggunakan *K-Means Clustering*

Hilma Fitri Solehah<sup>1</sup>, Ati Zaidiah<sup>2</sup>, Ika Nurlaili Isnainiyah<sup>3</sup>  
Sistem Informasi / Fakultas Ilmu Komputer  
Universitas Pembangunan Nasional Veteran Jakarta  
Jl. Rs. Fatmawati Raya, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450  
[hilmaafs28@gmail.com](mailto:hilmaafs28@gmail.com)<sup>1</sup>, [atizaidiah@upnvj.ac.id](mailto:atizaidiah@upnvj.ac.id)<sup>2</sup>, [nurlailiika@upnvj.ac.id](mailto:nurlailiika@upnvj.ac.id)<sup>3</sup>

**Abstrak.** Pandemi COVID-19 yang terjadi sejak awal tahun 2020 merupakan sebuah permasalahan dan tantangan baru yang dirasakan oleh dunia. Penularan infeksi virus SARS-CoV-2 sangat mudah dan terjadi dengan cepat. Salah satu faktor risiko yang rentan terinfeksi COVID-19 adalah faktor penyakit penyerta/ komorbiditas yang ada pada masyarakat. Untuk dapat mengidentifikasi kelompok komorbid apa saja yang banyak ditemukan pada pasien COVID-19, penelitian ini melakukan kegiatan analisis dan pemodelan menggunakan algoritma data mining *K-Means Clustering* yang akan menghasilkan cluster pasien COVID-19. Adapun atribut data yang digunakan yakni nilai tengah usia, jenis komorbiditas, serta persentase penyakit melihat ukuran populasinya. Data yang akan diolah didapatkan dari sumber valid Covid Analytics yang berisikan rangkuman data riset medis pasien COVID-19 dari berbagai Negara. Cluster yang dihasilkan berjumlah 3 cluster dengan perbedaan derajat dari ringan hingga berat. Terdapat 321 data (39%) kelompok pasien COVID-19 yang diolah termasuk pada cluster pertama dengan rata-rata usia 48 tahun dan komorbid terbanyak ditemukan yaitu penyakit diabetes. Cluster kedua terdiri dari 370 data (44%) dengan rata-rata usia 64 tahun dan penyakit diabetes merupakan komorbid terbanyak. Sedangkan untuk cluster ketiga terdiri dari 140 data (17%) dengan rata-rata usia 68 tahun dan komorbid terbanyak yaitu penyakit Hipertensi. Analisis menggunakan *K-Means Clustering* ini dapat memberikan gambaran kelompok pasien dengan derajat yang berbeda serta bentuk penanganan yang akan berbeda pula.

**Kata Kunci:** Covid-19, Komorbiditas, *K-Means Clustering*

### 1 Pendahuluan

COVID-19 (*Coronavirus Disease 2019*) adalah penyakit menular yang disebabkan oleh virus SARS-CoV2 (*Severe Acute Respiratory Syndrome Coronavirus 2*) yang dapat menyerang sistem pernapasan manusia. Menurut Herick et al. (2019), virus SARS-CoV2 tersebar melalui droplet respiratorik yang terjadi saat bersin dan batuk sehingga secara tidak langsung dapat ditemukan pada benda dan permukaan yang terkontaminasi [1]. Pada penelitian yang dilakukan Tiodora (2020) dengan menggunakan metode *Discourse Network Analysis*, ia sampai pada kesimpulan bahwa kelompok yang rentan terpapar COVID-19 yaitu kelompok lansia, penderita penyakit kronis, perokok, pria, dan orang dengan golongan darah A [2].

Studi yang dilakukan ini akan berfokus pada salah satu kelompok berisiko yaitu orang dengan penyakit kronis atau penyakit penyerta. Infeksi COVID-19 berat pada pasien dengan penyakit penyerta dapat mengalami risiko komplikasi kardiovaskular yang lebih tinggi [1]. Penelitian ini akan menggunakan informasi prevalensi komorbiditas pasien COVID-19 serta nilai tengah usia pada sebuah populasi untuk menggolongkan kelompok pasien berkomorbid dengan tiga kategori, yaitu berderajat ringan, berderajat sedang, dan berderajat berat.

Pada penelitian ini akan diimplementasikan metode *K-Means Clustering* dengan memanfaatkan pengetahuan *Machine learning* melalui tahapan analisis data, penentuan pola dan pembuatan model pengujiannya. Algoritma *K-Means Clustering* cukup efisien dan mudah diterapkan untuk pengklasteran sebuah dataset yang besar karena kompleksitas perhitungannya yang bersifat linier.

## 2 Tinjauan Pustaka

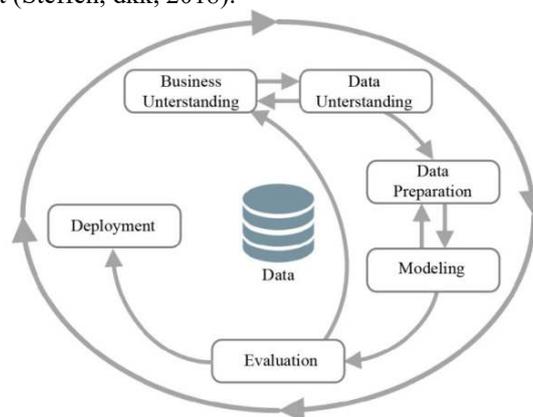
### 2.1 COVID-19 dan Komorbiditas

Pada awalnya penyakit ini dikenal dengan nama *Novel Coronavirus 2019 (2019-nCoV)* saat pertama kali dilaporkan pemerintah Tiongkok dengan kasus pertamanya di Wuhan, Provinsi Hubei, Tiongkok. Pada tanggal 11 Februari 2020, *World Health Organization (WHO)* memberi nama penyakit tersebut dengan *Coronavirus Disease 2019 (COVID-19)*. Melihat penyebaran virus yang cepat dan meluas, pada tanggal 11 Maret 2020 WHO menetapkan COVID-19 sebagai *Public Health Emergency of International Concern (PHEIC)* atau Kedaruratan Kesehatan Masyarakat yang Meresahkan Dunia (KKMMD), [3].

Faktor komorbiditas adalah penyebab pemberat penyakit yang dapat menyebabkan risiko yang dialami seseorang semakin tinggi [1]. COVID-19 dengan komorbiditas termasuk wanita hamil, bayi baru lahir, orang tua, dan pasien dengan komorbid seperti hipertensi, diabetes, dan penyakit kardiovaskular lainnya cenderung mengalami kejadian lebih parah dan memerlukan perawatan lebih ke unit perawatan intensif karena risiko yang dialaminya.

### 2.2 Metode CRISP-DM

Metode CRISP-DM (*Cross Industry Standard Processm For Data Mining*) merupakan sebuah standarisasi dalam pemrosesan data mining sebagai strategi dalam pemecahan masalah secara umum pada suatu bisnis atau unit perusahaan [4]. Metode ini digunakan untuk menerjemahkan masalah bisnis ke dalam metode *data mining* untuk melakukan pengembangan data yang sesuai. Terdapat enam fase dalam CRISP-DM yakni pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyebaran dengan alur proses yang dijelaskan pada gambar berikut (Steffen, dkk, 2018):



**Gambar. 1.** Fase dalam metode CRISP-DM yang membantu dalam pengolahan data dengan *data mining*.

### 2.3 Data Mining

Data mining adalah kegiatan mengekstrak informasi atau pengetahuan (*knowledge*) penting dari suatu set data berukuran besar dengan menggunakan teknik tertentu. Informasi yang dihasilkan dari data mining dapat digunakan untuk menemukan pola yang berguna ke depannya. Data mining melibatkan penggunaan metode atau alat untuk mendeteksi pola dan melakukan tugas prediksi. Ketepatan dari prediksi inilah keuntungan yang diharapkan dalam bisnis dari penggunaan data mining. Tugas-tugas yang biasa dilakukan oleh data mining antara lain adalah klastering, klasifikasi, regresi/estimasi, dan asosiasi [5].

### 2.4 K-Means Clustering

Dalam *K-Means Clustering*, kumpulan data  $s$  dibagi menjadi  $k$  kluster untuk meminimalkan jumlah jarak Euclidean antara masing-masing titik data dan pusat kluster terdekatnya. Implementasi *K-Means Clustering*

# KLASTERISASI PASIEN COVID-19 BERDASARKAN KOMORBIDITAS MENGGUNAKAN *K-MEANS CLUSTERING*

melibatkan optimasi untuk mencari alokasi kluster terbaik yang bisa meminimalkan jarak Euclidean dengan tahapan sebagai berikut [6]:

1. Menentukan jumlah cluster  $k$ , terdapat berbagai cara yang dapat dilakukan yakni memilih secara random atau menggunakan metode Elbow untuk mencari  $k$  yang optimal.
2. Menentukan centroid (titik pusat awal), dalam penentuannya dapat dilakukan secara acak sebanyak jumlah cluster  $k$  yang sudah ditentukan.
3. Mengalokasikan setiap data ke dalam cluster terdekat. Kedekatan data/obyek dapat menggunakan perhitungan *Euclidean Distance* dengan rumus:

$$D(X_j - C_j) = \sqrt{\sum_{j=0}^n (X_j - C_j)^2} \quad (1)$$

Keterangan:

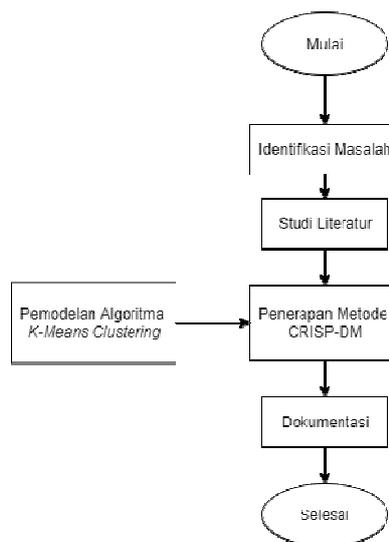
$X_j$  = obyek  $x$  ke- $j$

$C_j$  = pusat kluster  $j$ , atau *Centroid*

$n$  = banyaknya obyek atau anggota dalam cluster  $j$

4. Menghitung kembali pusat cluster dari kelompok cluster dan mengelompokkannya berdasarkan jarak terkecil di antara  $k$  cluster.
5. Ulangi langkah 3 dan 4 dengan *centroid* yang baru. Jika dalam perhitungannya *centroid* sudah tidak berubah, maka proses pengklasteran dapat selesai.

## 3 Metode Penelitian



**Gambar. 2.** Alur penelitian dalam melakukan pemrosesan *data mining*

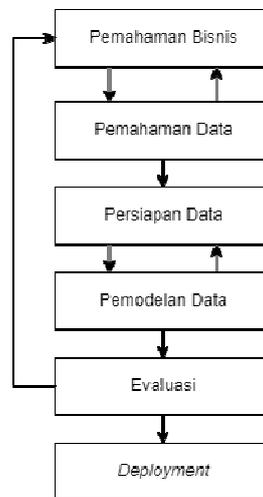
### 3.1 Identifikasi Masalah

Tahapan ini melakukan pengidentifikasian masalah untuk melakukan penelitian dengan mengimplementasikan model algoritma *K-Means Clustering* pada kelompok pasien COVID-19.

### 3.2 Studi Literatur

Studi literatur pada penyusunan tugas akhir ini berkaitan dengan teori pada buku, jurnal, dan studi penelitian mengenai metode yang digunakan yaitu klastering sebuah penyakit, algoritma *Data Mining*, *K-Means Clustering*. Dilakukan juga studi untuk memahami alat bantu yang akan digunakan untuk mengolah data.

### 3.3 Penerapan Metode CRISP-DM



**Gambar. 3.** Penerapan Metode CRISP-DM

### 3.4 Dokumentasi

Tahapan ini berupa melakukan dokumentasi sistem dan penarikan kesimpulan sebagai lanjutan dari fase terakhir pada metode CRISP-DM sebelumnya. Setelah dilakukan dokumentasi atas sistem dan penyelesaian laporan, maka analisis dan perbandingan kelompok pasien COVID-19 pada penelitian ini telah selesai.

## 4 Hasil dan Pembahasan

Dalam melakukan analisis dan perbandingan kelompok pasien COVID-19, diperoleh data yang bersumber dari situs [covidanalytics.io](https://www.covidanalytics.io) untuk mendukung proses klasterisasi sehingga dapat dibangun sebuah sistem yang menampilkan hasilnya. Penerapan *K-Means Clustering* dilakukan dengan beberapa tahapan untuk mendukung perancangan sistem dengan metode CRISP-DM

### 4.1 Pemahaman Bisnis

Covid Analytics adalah kelompok peneliti yang berkolaborasi dengan beberapa rumah sakit di Amerika Serikat untuk mengembangkan pengetahuan mengenai COVID-19. Hasil penelitian yang dilakukan kelompok riset tersebut diunggah secara publik melalui situs [www.Covidanalytics.io](https://www.Covidanalytics.io). Covid Analytics juga memberikan dataset hasil *review* berbagai jurnal kesehatan melihat studi klinis pasien COVID-19 dari berbagai jurnal. Karena berfokus kepada faktor komorbiditas dari pasien COVID-19, penentuan faktor dan pengidentifikasian batasan atau strategi awal pada penelitian ini di antaranya ditentukan sebagai berikut.

- a. Mendapatkan pengetahuan mengenai kelompok rentan.
- b. Menjadikan sebuah tolak ukur melihat pengaruh komorbid.
- c. Memanfaatkan pengetahuan bidang ilmu computer untuk menerapkan pemodelan pada data pasien COVID-19.
- d. Berfokus pada faktor nilai median usia, ukuran populasi, dan jenis komorbiditas pasien COVID-19.

### 4.2 Pemahaman Data

Data yang akan digunakan dalam penelitian ini adalah data kelompok pasien COVID-19 dari berbagai negara yang bersumber dari *paper* klinis kesehatan yang telah dilakukan *review* oleh tim Covid Analytics [7]. Sampel data kelompok pasien COVID-19 yang akan digunakan dapat dilihat pada tabel di bawah ini.

KLASTERISASI PASIEN COVID-19 BERDASARKAN  
KOMORBIDITAS MENGGUNAKAN *K-MEANS CLUSTERING*

**Tabel 1.** Sampel Dataset (Diakses 27 Januari 2022)

ID	SUB_ID	Country	Province	Study Pop Size	...	Projected Mortality
1.0	0.0	China	Wuhan	191.0	...	0.28
1.0	1.0	China	Wuhan	54.0	...	1
...	...	...	...	...	...	...
202.0	5.0	USA	Georgia	50.0	...	0.25

(Sumber: <https://www.covidanalytics.io/dataset>)

Jumlah dataset yang diperoleh yaitu 538 data dengan 162 variabel. Data tersebut masih terdapat kekurangan seperti adanya *missing values* dan data *outliers* yang menyebabkan dataset tersebut menjadi *bias* karena nilai data yang berbeda jauh.

### 4.3 Data Preprocessing

Untuk melakukan pengolahan dari dataset tersebut, terdapat atribut data yang dihilangkan dan penelitian ini akan menggunakan 10 atribut yaitu ID, *Country*, *Study Pop Size* (N), *Median Age*, *Hypertension*, *Diabetes*, *Cardiovascular Disease* (incl. CAD), *Chronic obstructive lung* (COPD), dan *Chronic kidney/renal disease*. Setelah itu, dilakukan pembersihan data dengan menghapus atribut yang tidak diperlukan, menghapus missing value, menghilangkan *outliers* dan ketidak-konsistenan data. Setelah data sudah bersih, akan dilakukan transformasi data untuk memudahkan dalam menerapkan model *clustering*-nya.

Atribut data yang akan dilakukan transformasi dengan inialisasi kategori dan proses *encoding* yang pertama dilakukan transformasi adalah kolom kelima penyakit komorbid yang diubah menjadi satu kolom dengan nama "*Comorbidities*" yang pada kolom tersebut tiap nama komorbid akan di *encoding* dengan kode 1-5. Sedangkan untuk nilai dari atribut komorbid sebelumnya akan dibuat pada atribut baru bernama "*Percentage*". Selanjutnya atribut nama Negara ditransformasikan dengan kode 0-11 setelah dilakukan *encoding* terhadap 12 data nama sumber Negara. Transformasi data terakhir yaitu pada atribut jumlah sampel/populasi (*Study Pop Size*) dengan inialisasi berdasarkan *range* ukuran. Untuk lebih lengkapnya dapat dilihat dari tabel transformasi data di bawah ini.

**Tabel 2.** Data Awal Penelitian setelah transformasi

No	ID	<i>Country</i>	N	<i>Median Age</i>	<i>Comorbidities</i>	<i>Percentage</i>
1	1	0	2	56	4	30
2	1	0	1	69	4	48,15
3	1	0	2	52	4	23,36
4	3	0	1	50	4	15
...	..	...	...	...	...	...
830	202	11	2	60	1	9,7
831	202	11	1	64,5	1	16

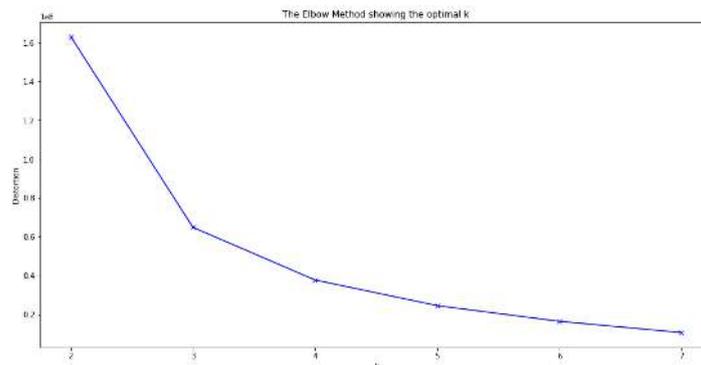
## KLASTERISASI PASIEN COVID-19 BERDASARKAN KOMORBIDITAS MENGGUNAKAN *K-MEANS CLUSTERING*

### 4.4 Proses Pemodelan *K-Means Clustering*

Pada tahapan ini akan dilakukan penerapan model *clustering* menggunakan *K-Means*. Tahapan proses *clustering* terhadap data penelitian yang akan dilakukan yaitu:

#### 4.4.1 Menghitung Jumlah Cluster $k$

Jumlah cluster  $k$  yang akan digunakan dalam penelitian ini ditentukan dengan metode *elbow*. Pada data awal penelitian ini akan dilakukan perhitungan SSE (*Sum Squared-error*) dengan pengujian jumlah cluster  $k = 2$  sampai dengan  $k = 8$ . Uji coba akan dilakukan dengan bahasa python dengan menghasilkan grafik mengenai nilai error yang dihasilkan. Berikut merupakan grafik yang merepresentasikan nilai SSE terhadap setiap  $k$ .



**Gambar. 4.** Grafik metode *elbow*. Pada sumbu  $y$  grafik tersebut merupakan *Distortions* atau nilai penjumlahan dari seluruh jarak *cluster*.

#### 4.4.2 Menentukan Centroid

Penentuan *centroid* atau titik pusat awal yang menjadi acuan untuk perhitungan *cluster* akan dilakukan secara acak dari data penelitian. *Centroid* dari masing-masing atribut data yang telah ditentukan. Adapun centroid yang digunakan pada atribut Median Age dan Percentage untuk C1, C2, dan C3 yakni: [(37), (8.3)], [(56), (30)], [(83), (40.6)]

#### 4.4.3 Menghitung jarak setiap data ke pusat cluster

Tahapan selanjutnya menentukan jarak terdekat antara *cluster* dengan tiap data menggunakan perhitungan jarak dengan rumus *Euclidean Distance*. Setelah dilakukan perhitungan pada seluruh data, hasil perhitungan tiap data dengan ketiga *centroid* untuk iterasi pertama dan dilanjutkan dengan perhitungan jarak terkecil antara data dengan *centroid* untuk menentukan hasil *centroid*-nya.

#### 4.4.4 Menghitung Centroid Baru

Tahapan selanjutnya dilakukan perhitungan *centroid* baru dengan menghitung rata-rata dari tiap clusternya.

#### 4.4.5 Melakukan iterasi dengan centroid baru sampai clustr tidak berubah

Proses perhitungan seperti tahapan kedua dilakukan secara terus menerus jika masih ditemukan perubahan pada *centroid*. Proses iterasi akan berhenti saat nilai centroid terakhir tidak berubah. Pada perhitungan sebagian datayang dilakukan secara manual, nilai centroid tetap berakhir pada iterasi ke-11.

KLASTERISASI PASIEN COVID-19 BERDASARKAN KOMORBIDITAS MENGGUNAKAN *K-MEANS CLUSTERING*

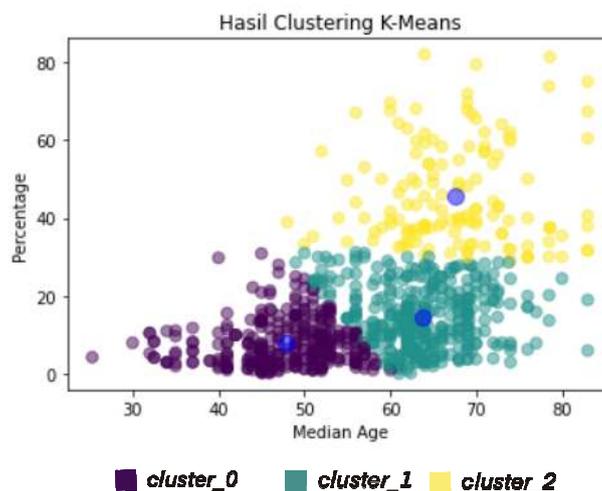
4.4.6 Hasil Perhitungan Centroid Terakhir

Tabel 3. Hasil Centroid Terakhir

A (Median Age)	B (Percentage)	
47.77545455	7.88884013	C1 (cluster_0)
63.78215054	14.59943548	C2 (cluster_1)
67.53207143	45.69642857	C3 (cluster_2)

Hasil iterasi terakhir tersebut diperoleh dengan bahasa python untuk menentukan centroid pada hasil clustering. Maka proses clustering pun berakhir dengan perolehan Centroid Cluster 1: (47,780 ; 7,890), Centroid Cluster 2 : (63,783; 14,600 ), Centroid Cluster 3 : ( 67,530; 45,700 ). Jumlah data dari hasil pemodelan K-Means Clustering memperoleh hasil sebanyak 140 data untuk cluster pertama, 321 data untuk cluster kedua, dan 370 data untuk cluster ketiga.

Dari hasil analisis tiap *cluster* yang tersebut, didapatkan penempatan tiap cluster dan letak *centroid* nya berdasarkan atribut persentase komorbid dan nilai median usia yang digambarkan pada grafik *plotting* sebagai berikut.



Gambar. 5. Plot Hasil Clustering. *Plot* digambarkan dalam bentuk diagram *scatter*.

4.5 Perhitungan Evaluasi

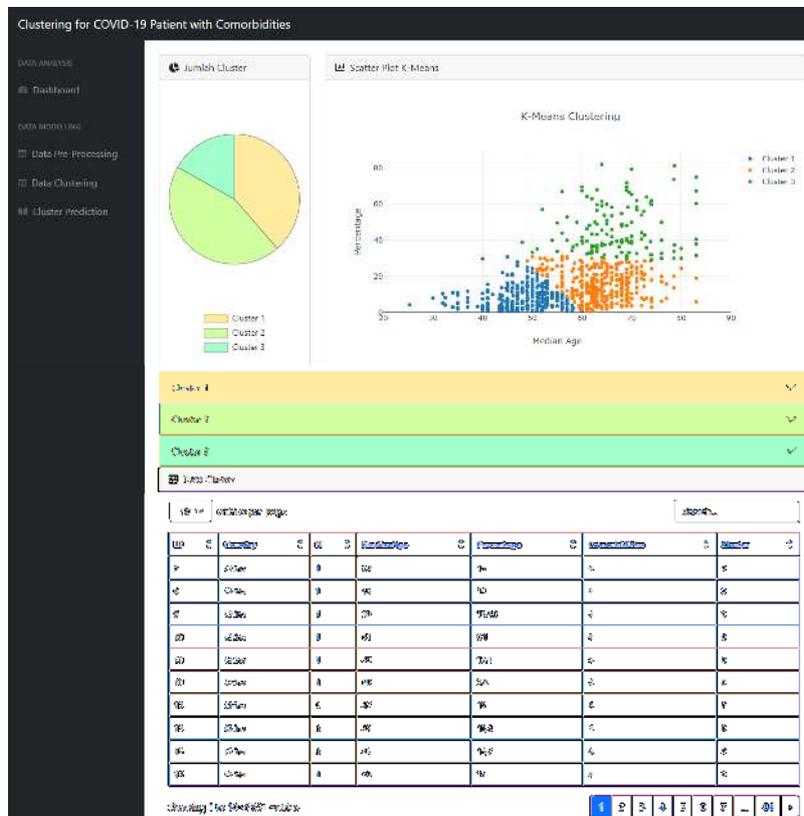
Tahapan evaluasi dilakukan pada hasil model cluster yang telah dibangun dengan menghitung SSE (*Sum of Squared-error*) dari nilai pada masing-masing *cluster* yang dihasilkan guna menguji validasinya. Dilakukan perhitungan nilai SSE dimulai dari nilai  $k=2$  sampai  $k=8$ . Semakin kecil nilai SSE pada suatu *cluster*, maka itulah *cluster* terbaik yang dihasilkan. Pada penelitian ini, didapatkan hasil perhitungan SSE yang paling kecil yakni pada  $k=3$  dengan nilai  $6,487165e+16$  (64871648.882278666).

4.6 Deployment

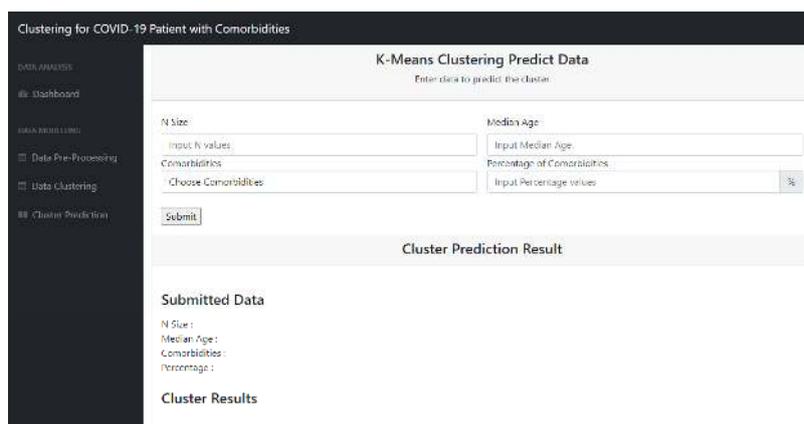
Pada tahapan deployment dilakukan pembangunan sistem sederhana dengan algoritma k-means clustering untuk memprediksi cluster berdasarkan hasil pemodelan. Pada halaman data clustering, menampilkan hasil dari proses K-Means Clustering yang membentuk 3 buah cluster dengan grafik *pie chart* untuk menunjukkan jumlah data yang terbagi pada tiap *cluster* dan grafik *scatter plot* yang menampilkan hasil *plotting* tiap *cluster*. Halaman data clustering juga menampilkan grafik data perolehan tiap *cluster* untuk atribut data *Percentage*, *Comorbidities*, dan *Median Age*. Bagian bawah juga menampilkan tabel hasil clustering dengan kategori 0 untuk cluster 1, nilai

## KLASTERISASI PASIEN COVID-19 BERDASARKAN KOMORBIDITAS MENGGUNAKAN *K-MEANS CLUSTERING*

1 untuk cluster 2, dan nilai 2 untuk cluster 3. Sedangkan pada halaman prediksi clustering, terdapat form input atribut *Median Age* dan *Percentage*, serta memuat pemrosesan prediksi dari data hasil *clustering* K-Means. Berikut merupakan tampilan dari sistem yang dibangun.



**Gambar. 6.** Halaman Data Clustering.



**Gambar. 7.** Halaman Prediksi Clustering.

## 5 Kesimpulan

Berdasarkan proses klasterisasi dengan algoritma K-Means Clustering, didapatkan kesimpulan sebagai berikut dengan melakukan tinjauan pustaka beberapa jurnal kesehatan dan pedoman tata laksana COVID-19 [8] sebagai berikut.

## KLASTERISASI PASIEN COVID-19 BERDASARKAN KOMORBIDITAS MENGGUNAKAN K-MEANS CLUSTERING

1. Kelompok *cluster* 1 dapat dikatakan sebagai kelompok dengan derajat klinis berkomorbid ringan dengan persentase komorbid di bawah 30% dari 321 data sampel bervariasi. Pasien COVID-19 berkomorbid yang memiliki rata-rata usia 48 tahun menurut Pedoman Tata laksana COVID-19 melihat derajatnya dapat melakukan isolasi maksimal 10 hari sejak onset pertama kali hingga 3 hari bebas gejala. Untuk pasien COVID-19 yang memiliki penyakit penyerta dianjurkan untuk melakukan isolasi di fasilitas kesehatan supaya pengobatan komorbid tetap diperhatikan sehingga tidak terjadi komplikasi tanpa pemantauan medis.
2. Kelompok *cluster* 2 merupakan kelompok dengan derajat sedang yang memiliki persentase berkomorbid di bawah 30% dan rata-rata usia 64 tahun. Menurut Pedoman Tata laksana COVID-19, pasien dengan derajat sedang dapat melakukan isolasi di ruang perawatan rumah sakit darurat COVID-19 karena mendapatkan pemantauan secara berkala oleh tenaga medis.
3. Kelompok *cluster* 3, termasuk pada derajat berat-kritis karena mayoritas berusia di atas 50 tahun dengan rata-rata yang ditemukan pada cluster ini adalah 68 tahun. Persentase dari cluster ini juga di atas 30% melihat 140 data yang termasuk ke dalam cluster tersebut. Penanganan bagi pasien berderajat berat ini dilakukan pemantauan dan isolasi di ruang isolasi Rumah Sakit Rujukan atau rawat secara kohorting dengan pemantauan yang kompleks.

## Referensi

- [1] Willim, H. A., Ketaren, I. & Supit, A. I., "Dampak Coronavirus Disease 2019 terhadap Sistem Kardiovaskular". *Jurnal e-Clinic (eCl)*. Vol. VIII(2), pp.237-245. 2019.
- [2] Siagian, T. H., "Mencari Kelompok Berisiko Tinggi Terinfeksi Virus Corona dengan Discourse Network Analysis", *Jurnal Kebijakan Kesehatan Indonesia: JKKI*. Vol.IX(2), pp. 98-106. 2020.
- [3] Kemendagri, Gugus Tugas COVID-19. (2020). Pedoman Umum Menghadapi Pandemi COVID-19 Bagi Pemerintah Daerah, Pencegahan, Pengendalian, Diagnosa, Manajemen. [Online] Available at: <https://covid19.go.id/p/panduan/kemendagri-pedoman-umum-menghadapi-pandemi-covid-19-bagi-pemerintah-daerah>.
- [4] Huber, S., Wiemer, H., Schneider, D. & Ihlenfeldt, S., "DMME: Data mining methodology for engineering applications- a holistic extension to the CRISP-DM model", *12<sup>th</sup> CIRP Conference on Intelligent Computation in Manufacturing Engineering*, Vol. LXXIX, pp.403-408. 2018.
- [5] Indraputra, R. A., Fitriana, R., "K-Means Clustering Data COVID-19" *Jurnal Teknik Industri*. X(3), pp. 275-282. 2020.
- [6] Santosa, B. & Umam, A., 2018. *Data Mining dan Big Data Analytics*. Yogyakarta: Penebar Media Pustaka.
- [7] Bertsimas, Dimitris and Bandi, Hari and Boussioux, Leonard and Cory-Wright, Ryan and Delarue, Arthur and Digalakis, Vasileios and Gilmour, Samuel and Graham, Justin and Kim, Adam and Lahlou Kitane, Driss and Lin, Zhen and Lukin, Galit and Li, Michael and Mingardi, Luca and Na, Liangyuan and Orfanoudaki, Agni and Papalexopoulos, Theodore and Paskov, Ivan and Pauphilet, Jean and Skali Lami, Omar and Sobiesk, Matthew and Stellato, Bartolomeo and Carballo, Kimberly and Wang, Yuchen and Wiberg, Holly and Zeng, Cynthia. (2020). *An Aggregated Dataset of Clinical Outcomes for COVID-19 Patients* [Online]. Available at: [http://www.covidanalytics.io/dataset\\_documentation](http://www.covidanalytics.io/dataset_documentation).
- [8] PDPI, PERKI, PAPDI, PERDATIN, IDAI., (2020). Pedoman Tata laksana COVID-19 Edisi Ketiga. [Online] Available at: <https://www.papdi.or.id/pdfs/983/Buku%20Pedoman%20Tatalaksana%20COVID-19%205OP%20Edisi%203%202020.pdf>
- [9] Drew, S., Adisasmita, A. C., "Gejala dan komorbid yang memengaruhi mortalitas pasien positif COVID-19 di Jakarta Timur, Maret-September 2020", *Tarumanagara Medical Journal*. Vol.III(3), pp. 274-283. 2021.
- [10] Ramdani, A. L. & Firmansyah, H. B., "Clustering Application for UKT Determination Using Pillar K-Means Clustering Algorithm and Flask Web Framework", *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIM)*. Vol I(2), pp. 53-59. 2018.
- [11] RSUP dr. Soeradji Tirtonegoro, (2020). *Penyakit Kardiovaskular dan Pandemi COVID-19*. [Online] Available at: <https://rsupsoeradji.id/penyakit-kardiovaskular-dan-pandemi-covid-19/>