

Optimasi Nilai K Pada Algoritma K-Means Untuk Klasterisasi Data Pasien Covid-19

Moh. Fatkuroji¹, Fajrizal², Taslim³, Eka Sabna⁴, Kursiah Warti Ningsih⁵

rojiunilak@gmail.com¹, fajrizal@unilak.ac.id², taslim@unilak.ac.id³, es3jelita@yahoo.com⁴,

kursiahwarti@gmail.com⁵

^{1,2,3}Universitas Lancang Kuning

⁴Universitas Hang Tuah

⁵STIKes Payung Negeri

Informasi Artikel

Diterima : 21 Aug 2022

Direview : 28 Aug 2022

Disetujui : 30 Aug 2022

Kata Kunci

Covid-19, *Clustering*, *K-Means*, *Elbow*, *Python*.

Abstrak

Dengan menyebarluasnya Covid-19 hingga di berbagai negara menyebabkan Pemerintah dan Instansi Kesehatan di dunia kesulitan dalam menangani kasus Covid-19 hingga sampai saat ini. Adapun pencegahan yang dilakukan oleh Pemerintah maupun Instansi Kesehatan di dunia dilakukan dengan cara pemberian vaksin kepada masyarakat. Akan tetapi pada beberapa tempat tidak terlaksana sesuai dengan PMK Nomor 84 Tahun 2020 yang mengutamakan pemberian vaksin pada masyarakat lansia. Dengan padatnya penduduk di Indonesia saat ini menyebabkan pemberian vaksin tidak melihat siapa yang lebih dahulu diprioritaskan. Penerapan algoritma *k-means* dilakukan untuk mengklusterkan pasien yang terkena Covid-19 pada data kasus Covid-19 yang didapat dari [kaggle.com](https://www.kaggle.com) berupa data pasien sejak 01 Januari 2020 sampai dengan 31 Mei 2020 sebanyak 139.119 kasus. Hasil *clustering* data kasus yang terkena Covid-19 dengan $k=3$ menghasilkan nilai WCSS sebesar 6801292.2. Perhitungan Algoritma K-Means dengan bantuan Tools Google Colaboratory menghasilkan cluster dengan kasus pasien terkena Covid-19 pada *Cluster-0* sebanyak 58.237 kasus, pada *Cluster-1* sebanyak 53.932 kasus, dan pada *Cluster-2* sebanyak 26.950 kasus.

Keywords

Covid-19, *Clustering*, *K-Means*, *Elbow*, *Python*.

Abstrak

With the spread of Covid-19 to various countries, it is difficult for Governments and Health Agencies in the world to handle Covid-19 cases to date. The prevention carried out by the Government and Health Agencies in the world is carried out by giving vaccines to the public. However, in some places it is not implemented in accordance with PMK Number 84 of 2020 which prioritizes providing vaccines to the elderly. With the current density of the population in Indonesia, the administration of vaccines does not see who is prioritized first. The application of the k-means algorithm is carried out to cluster patients affected by Covid-19 on the Covid-19 case data obtained from [kaggle.com](https://www.kaggle.com) in the form of patient data from January 1, 2020 to May 31, 2020 as many as 139119 cases. The results of clustering data on cases affected by Covid-19 with $k=3$ yielded a WCSS value of 6801292.2. Calculations of the K-Means Algorithm using the Google Collaboratory Tools resulted in clusters with the cases of patients affected by Covid-19 in Cluster-0 as many as 58.237 cases, in Cluster-1 as many as 53.932 cases, and in Cluster-2 as many as 26.950 cases.

A. Pendahuluan

Pada pertengahan bulan Desember 2019, kasus penyakit gangguan saluran pernapasan misterius muncul di Wuhan, Provinsi Hubei pada pertama kalinya. Kasus penularan penyakit ini pada awalnya belum diketahui berasal dari mana, tetapi kejadian pertama kalinya dihubungkan pada pasar tradisional di Wuhan. Pada 18 Desember sampai 29 Desember 2019, terdapat sebanyak lima orang terkena gangguan pernapasan berat atau ARDS (*Acute Respiratory Distress Syndrome*) yang dirawat. Mulai dari 31 Desember 2019 sampai 3 Januari 2020 jumlah kasus melonjak tinggi, diawali dengan adanya laporan sebanyak 44 kasus. Mendekati jangka waktu sebulan, penyakit ini sudah meluas ke berbagai daerah dan provinsi lain di China dan negara besar sekitaran benua Asia. Covid-19 pada awalnya di Indonesia muncul laporan sebanyak dua kasus pada tanggal 2 Maret 2020. Pada data 31 Maret 2020, kasus meningkat dan dipastikan dengan jumlah 1.528 kasus serta kasus pasien meninggal dengan total 136 orang. Angka kematian kasus Covid-19 di Indonesia sebesar 8,9%, merupakan angka yang paling besar di Asia Tenggara[1]. Dengan menyebarkan Covid-19 hingga diberbagai negara menyebabkan Pemerintah dan Instansi Kesehatan di dunia kesulitan dalam menangani kasus Covid-19 hingga sampai saat ini. Adapun pencegahan yang dilakukan oleh Pemerintah maupun Instansi Kesehatan di dunia dilakukan dengan cara pemberian vaksin kepada masyarakat. Mengutip Peraturan Menteri Kesehatan Nomor 84 Tahun 2020 tentang Pelaksanaan Kegiatan Vaksinasi Dalam Bentuk Pencegahan atau Penanganan *Corona Virus Disease* 2019 (Covid-19) menjelaskan tentang Pelaksanaan dan Pelayanan Kegiatan Vaksinasi Covid-19 bahwasanya jadwal serta tahapan dalam pemberian Vaksin Covid-19 sudah ditetapkan sesuai dengan adanya Vaksin Covid-19 yang tersedia, warga penerima Vaksin Covid-19 yang diutamakan dan juga ragam Vaksin Covid-19[2]. Akan tetapi pada beberapa tempat tidak terlaksana sesuai dengan peraturan tersebut. Pemberian vaksin selama ini tidak melihat siapa yang lebih dahulu diprioritaskan, sehingga bagi yang seharusnya diprioritaskan tidak mendapatkan vaksin.

Berdasarkan permasalahan tersebut, tidak menutup kemungkinan perlunya suatu upaya penanganan dalam mencari tahu informasi terbaru. Upaya tersebut salah satunya yaitu, dengan mengelompokkan pasien yang terdampak Covid-19. Pengelompokan ini bertujuan untuk mengetahui informasi baru dengan menggunakan teknik *Datamining Clustering* pada pasien yang terkena Covid-19 dan terkait dengan beberapa penyakit yang berkaitan dengan pernapasan ataupun organ tubuh paru-paru yang tidak menutup kemungkinan mempengaruhi terjangkitnya Virus Covid-19, seperti *Pneumonia* (Radang Paru-paru), *COPD* (Penyakit Paru Obstruktif Kronik), *Asthma* (Penyakit Asma), dan *INMSUPR* (Kekurangan Imun Dalam Tubuh) serta data pendukung seperti *Age* (Umur), *Contact Other Covid* (Melakukan Hubungan dengan Pasien yang Terserang Covid-19), dan *Covid Result* (Hasil Covid). Pada Penelitian ini, hasil penelitiannya diharapkan dapat dijadikan bahan evaluasi ataupun rekomendasi bagi Pemerintah dan Instansi Kesehatan dalam meningkatkan penanganan kasus agar lebih efektif terhadap pandemi Covid-19. Pemerintah dan Instansi Kesehatan nantinya dapat lebih memperhatikan terkait pengelompokan pasien berdasarkan umur serta beberapa penyakit yang berkaitan dengan pernapasan dan organ tubuh paru-paru, sehingga penanganan kasus Covid-19 lebih mudah diatasi. *Data mining*

merupakan tahapan untuk mendapatkan bentuk atau juga informasi yang dapat dimanfaatkan secara langsung berdasarkan sekelompok data dengan jumlah besar atau banyak, seringkali data mining juga disebut sebagai salah satu dari *Knowledge Discovery in Database* (KDD) yang merupakan sebuah tahapan dalam mencari informasi yang dapat dimanfaatkan berdasarkan data yang ada[3]. *Clustering* atau juga disebut pengklasteran merupakan suatu teknik datamining yang digunakan dalam melakukan analisis data untuk mendapatkan informasi baru dalam pengelompokan data atau membagi dari dataset ke beberapa subset[4]. Dalam penerapannya metode *clustering* dihubungkan dengan tahapan algoritma agar tujuannya dapat tercapai. *K-Means* sendiri merupakan salah satu metode *clustering* berdasarkan nilai jarak yang memisahkan sekumpulan informasi ke beberapa jumlah klaster dengan tipe data *numeric*. Algoritma *K-Means* merupakan *partitioning clustering* yang membagi-bagi data ke k bagian posisi yang berbeda[5].

Penelitian terdahulu yang dilakukan oleh Winarta dan Kurniawan bertujuan untuk mendemonstrasikan kinerja *Elbow* dalam membuat klaster data penggunaan narkoba yang optimal menggunakan algoritma *K-Means* sebagai metode pengelompokan data. Optimalisasi *cluster* didapatkan dari metode *Elbow* yang dilakukan oleh Google Collaboratory dengan menggunakan bahasa pemrograman Python. Hasil pengujian menunjukkan bahwa metode *elbow* bekerja sangat baik untuk menghasilkan *cluster* yang optimal pada $k = 3$. Di sini selisih nilai SSE adalah 1257.862 dengan uji $k = 5$ [6].

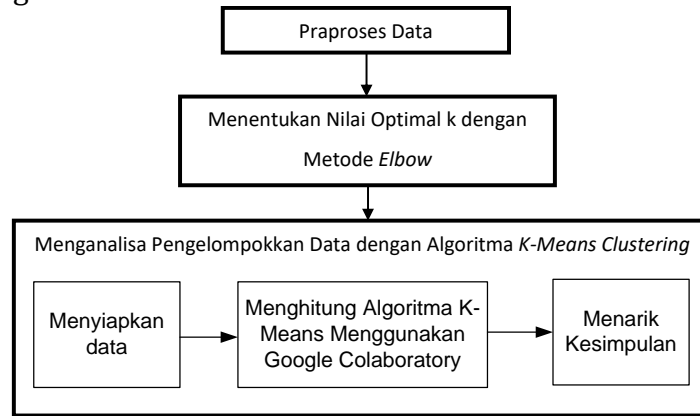
Penelitian sebelumnya juga melibatkan optimasi jumlah k menggunakan metode *Elbow* dimulai dengan mencari jumlah *cluster* terbaik menggunakan metode *Elbow*, dimana $K = 3$ merupakan jumlah *cluster* terbaik. Kemudian menggunakan teknik *clustering* untuk mengetahui langkah-langkah *clustering* dan pencarian dengan algoritma *K-Means*. $K = 3$ dikombinasikan dengan S (partikel acak) 5, 10, 15 untuk menguji. Hasil pengujian kemudian dievaluasi menggunakan SSE (*Sum of Square Error*) dan *Silhouette*. Grup terbaik dibuat dengan $K = 3$ dengan kombinasi $S = 10$. Nilai evaluasi yang dihasilkan oleh SSE adalah 2,4402 dan *Silhouette* adalah 0,5629, diperoleh Sari dkk menggunakan tools WEKA [7].

Karya Siagian et al sebelumnya bertujuan untuk menerapkan metode *clustering* menggunakan algoritma *K-Means* dan *K-Medoids*. Hasil penelitian menunjukkan bahwa algoritma *K-Means* dan *K-Medoids* sama-sama memberikan hasil *cluster* yang optimal $k = 3$. Hasil ini juga sesuai dengan hasil metode *Elbow* dan memeriksa validitas indeks *Davies Bouldin* untuk hasilnya. Jumlah *cluster* yang optimal adalah 3 (tiga) [8].

Berdasarkan referensi penelitian sebelumnya mengenai optimasi nilai k menggunakan metode *Elbow* pada algoritma *K-Means* menunjukkan hasil pencarian yang berbeda dan tidak selalu memberikan hasil yang baik dan efisien tergantung dari kondisi permasalahan dan datanya. Oleh karena itu, pada penelitian ini akan dilakukan *clustering* dengan algoritma *K-Means* untuk memberikan wawasan terkait *clustering* data pasien Covid-19 untuk mengenali karakteristik data pasien, skor yang lebih baik dan kelompok pasien.

B. Metode Penelitian

Beberapa tahapan penelitian ini merupakan gambaran atau penjelasan dari tahapan yang mesti dilakukan. Adapun tahapan penelitian yang harus dilakukan dijelaskan pada gambar berikut :



Gambar 1. Tahapan Penelitian

1) Praproses Data

Tahap pertama dari analisis adalah melakukan persiapan data. Data sebagai bahan perhitungan yang digunakan pada penelitian ini adalah data Covid-19 *patient pre-condition dataset* (data COVID-19 yang didapat dari pemerintah Mexico) yang diunduh dari situs kaggle.com. Dataset awal yang digunakan yaitu data kasus Covid-19 berupa data pasien sejak 01 Januari 2020 sampai dengan 31 Mei 2020 dengan total data pasien sebanyak 566.602 kasus. Adapun variabelnya terdapat pada gambar berikut ini :

Table 1. Informasi Variabel Data Covid-19

Array	Column	Non-Null Count	Dtype
0	id	566602 non-null	object
1	sex	566602 non-null	int64
2	patient_type	566602 non-null	int64
3	entry_date	566602 non-null	object
4	date_symptoms	566602 non-null	object
5	date_died	566602 non-null	object
6	intubed	566602 non-null	int64
7	pneumonia	566602 non-null	int64
8	age	566602 non-null	int64
9	pregnancy	566602 non-null	int64
10	diabetes	566602 non-null	int64
11	copd	566602 non-null	int64
12	asthma	566602 non-null	int64
13	inmsupr	566602 non-null	int64
14	hypertension	566602 non-null	int64
15	other_disease	566602 non-null	int64
16	cardiovascular	566602 non-null	int64
17	obesity	566602 non-null	int64
18	renal_chronic	566602 non-null	int64
19	tobacco	566602 non-null	int64
20	contact_other_covid	566602 non-null	int64
21	covid_res	566602 non-null	int64
22	icu	566602 non-null	int64

2) Optimasi Nilai K Dengan Metode *Elbow*

Elbow adalah sebuah suatu yang diterapkan dalam mencari suatu informasi untuk menetapkan jumlah k atau kluster terbaik berdasarkan hasil perbedaan antara jumlah k atau kluster yang dapat membentuk titik sudut siku. Dengan *Elbow* dapat menghasilkan ketetapan dengan cara menentukan nilai k atau kluster lalu menambah nilai k atau *cluster* tersebut agar dapat dijadikan bentuk data pada penentuan kluster terbaik[9].

Jumlah *cluster* k yang dihasilkan dari pengujian dengan *K-Means* dievaluasi menggunakan teknik SSE. Untuk menghitung SSE dapat menggunakan rumus sebagai berikut[6]:

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} ||x_i - y_k||_2^2$$

3) Pengelompokkan Data Dengan *K-Means Clustering*

K-Means adalah metode *clustering* berdasarkan jarak dengan memisahkan data ke sejumlah kelompok bagian atau kluster dengan tipe data *numeric*. *K-Means* juga merupakan *partitioning clustering* yang membagi data ke k bagian posisi yang terpisah[5]. Tahap yang harus dilalui dalam membentuk *clustering* dengan penerapan *K-Means* adalah[10]:

- Tetapkan jumlah *cluster* (k) berdasarkan hasil optimasi metode *Elbow*.
- Tetapkan k sebagai titik tengah kluster (*centroid*), dipilih dengan acak (*random*).
- Melakukan perhitungan jarak antara data pada masing-masing titik tengah kluster dengan perhitungan Euclidean Distance seperti pada rumus berikut[7]:

$$J(a_i, b_i) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$$

- Kelompokkan data berdasarkan nilai terdekat dengan centroid lalu lakukan pembaruan nilai titik pusat kluster baru berdasarkan posisi dari titik pusat kluster dengan persamaan berikut[10]:

$$w_k = \frac{1}{n_k} \sum j_i$$

4) Menarik Kesimpulan

Pada tahap ini menarik kesimpulan berdasarkan hasil pengelompokkan dengan menggunakan *K-Means Clustering*.

C. Hasil dan Pembahasan

Pengujian optimasi nilai k cluster serta perhitungan algoritma *K-Means Clustering* dilakukan dengan menggunakan bahasa pemrograman Python pada Google Colaboratory. Python merupakan bahasa pemrograman dalam bentuk skrip

(*scripting-language*) yang berpusat pada objek[11]. Adapun tahapannya adalah sebagai berikut :

1) Praproses Data

Sebelum data mining dapat dilakukan, perlu dilakukan pembersihan data sesuai dengan KDD. Proses pembersihan meliputi, antara lain, menghapus data duplikat, memeriksa ketidakkonsistenan data, dan memperbaiki kesalahan pada data, seperti salah ketik[12]. Tahapan pembersihan data dimaksudkan untuk menghilangkan nilai kosong atau tidak lengkap[13]. Pada tahap ini dilakukan pendeteksian *missing values* namun tidak ditemukan data yang hilang atau kosong lalu menghapus data yang tidak diperlukan seperti data id dan waktu pada dataset dengan rincian variabel seperti pada tabel 2 berikut :

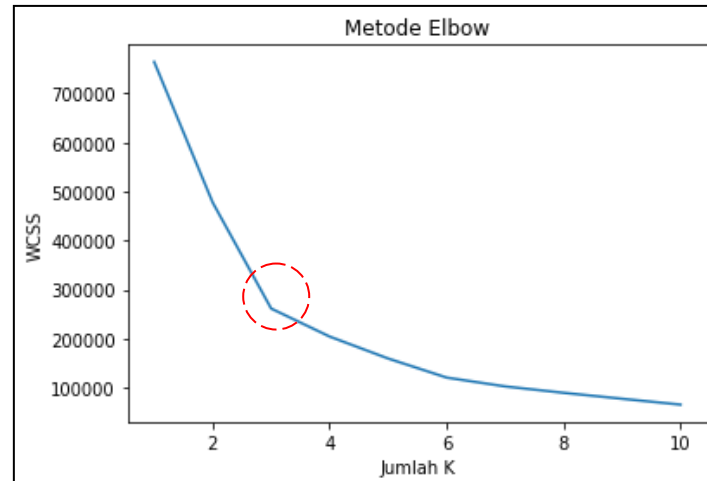
Table 2. Data Variabel

No	Nama Variabel	Tipe Data
1	sex	integer
2	patient_type	integer
3	intubed	integer
4	pneumonia	integer
5	age	integer
6	pregnancy	integer
7	diabetes	integer
8	copd	integer
9	asthma	integer
10	inmsupr	integer
11	hypertension	integer
12	other_disease	integer
13	cardiovascular	integer
14	obesity	integer
15	renal_chronic	integer
16	tobacco	integer
17	contact_other_covid	integer
18	covid_res	integer
19	icu	integer

Pada tabel diatas merupakan 19 data variabel yang digunakan sebagai bahan perhitungan *K-Means Clustering* dengan total 566.602 pasien covid-19 yang selanjutnya data tersebut di normalisasikan dengan *MinMaxScaler* pada Google colaboratory.

2) Optimasi Nilai K Dengan Metode *Elbow*

Menentukan nilai k optimal dilakukan dengan menerapkan Metode *Elbow*. Menurut Putu[9], jika nilai k atau *cluster* sebelumnya dengan nilai k atau *cluster* selanjutnya menampilkan titik sudut pada gambar grafik atau nilainya menunjukkan penurunan cukup tinggi maka k atau *cluster* tersebut yang terbaik dengan perhitungan *Within Cluster Sum of Square Error* (WCSS) yang adalah perbandingan antara jumlah kuadrat dan jumlah kuadrat yang mewakili jumlah data yang dikumpulkan per *centroid*[14]. Berikut visualisasi Metode *Elbow* berdasarkan perhitungan WCSS dengan *range* k atau *cluster* 1 sampai 10 dalam bentuk grafik :



Gambar 2. Grafik Optimasi Nilai K dengan Metode *Elbow*

Pada *K-Means Clustering*, nilai SSE diharapkan menjadi minimal dengan mencari jumlah *cluster* yang tepat dan titik tengah *cluster* yang tepat (*centroid*)[15]. Pada grafik diatas (Gambar 2) menampilkan hasil *clustering* data kasus yang terkena Covid-19 dengan $k=3$ menghasilkan nilai WCSS sebesar 261181.5 dan membentuk sudut siku-siku dan pengurangan nilai WCSS tidak terlalu besar seperti pada tabel 2 berikut :

Table 3. Hasil Perhitungan *Within Cluster Sum of Square Error* (WCSS)

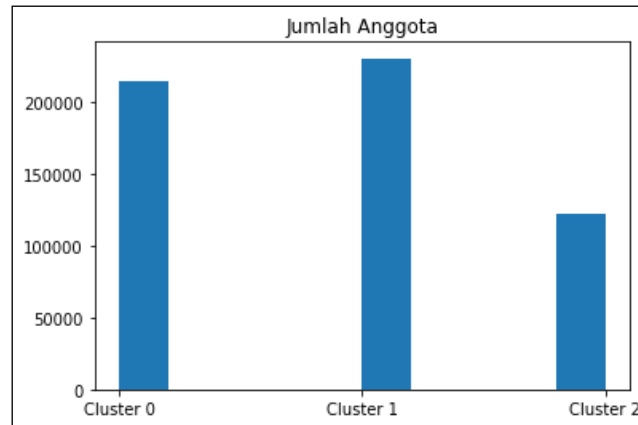
Jumlah k	Hasil WCSS
1	764288.8
2	477570.3
3	261181.5
4	204027.3
5	159236.2
6	120353.3
7	102322.3
8	89269.4
9	77131.5
10	65184.1

3) Pengelompokkan Data Dengan *K-Means Clustering*

Pengelompokkan ditentukan berdasarkan $k=3$ yang dihasilkan dari perhitungan SSE metode *Elbow*. Dengan menerapkan *library python* `sklearn.cluster KMeans` dan menghasilkan anggota tiap *cluster* sebagaimana pada tabel berikut :

Table 4. Anggota *Cluster*

Cluster	Total Anggota
0	214597
1	230216
2	121789



Gambar 3. Grafik Total Anggota Cluster

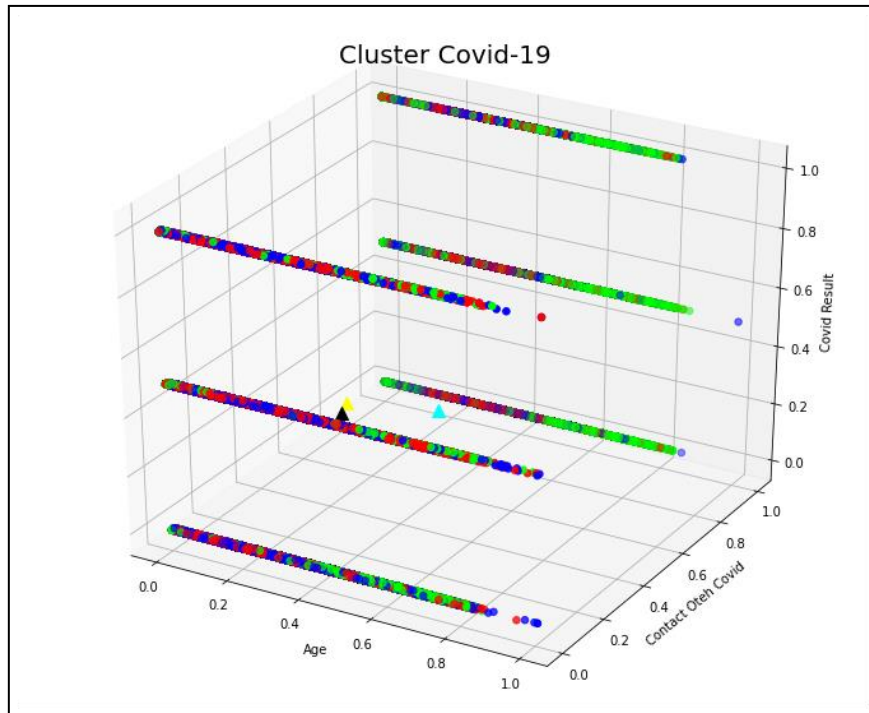
Pada Gambar 4 dapat dilihat jumlah pasien terkena Covid-19 pada *cluster-0* dengan total 214.597 kasus, pada *cluster-1* dengan total 230.216 kasus, dan pada *cluster-2* dengan total 121.789 kasus. Hasil yang dapat dilihat selanjutnya titik pusat *cluster* pada masing-masing variabel pada gambar berikut :

Tabel 1. Cluster Centers

No	Nama Variabel	Cluster 0	Cluster 1	Cluster 2
1	sex	9.987452e-01	5.149214e-13	5.993891e-01
2	patient_type	3.35849092e-04	2.25664082e-04	1.00000000e+00
3	intubed	9.79598691e-01	9.79596442e-01	9.36916490e-03
4	pneumonia	9.86447233e-03	9.92580983e-03	4.03005466e-03
5	age	3.32200659e-01	3.31018866e-01	4.41378258e-01
6	pregnancy	9.89703657e-01	1.51231173e-02	5.98598319e-01
7	diabetes	1.28044390e-02	1.20397425e-02	1.27453828e-02
8	copd	1.31898672e-02	1.24701327e-02	1.45902162e-02
9	asthma	1.29925125e-02	1.21592406e-02	1.47971822e-02
10	inmsupr	1.34511312e-02	1.27835749e-02	1.54273070e-02
11	hypertension	1.21773668e-02	1.14540180e-02	1.18466677e-02
12	other_disease	1.39565399e-02	1.31647522e-02	1.81523173e-02
13	cardiovascular	1.30940752e-02	1.24796174e-02	1.50255643e-02
14	obesity	1.17279330e-02	1.09102588e-02	1.33371617e-02
15	renal_chronic	1.31840004e-02	1.25360334e-02	1.46458303e-02
16	tobacco	1.24602225e-02	1.20862264e-02	1.45931789e-02
17	contact_other_covid	2.43397681e-01	2.69314494e-01	5.13381206e-01
18	covid_res	3.76880988e-01	3.97774171e-01	2.78937342e-01
19	icu	9.79598691e-01	9.79596442e-01	9.36497566e-03

4) Visualisasi Hasil Cluster

Pada Gambar berikut ini merupakan gambaran sebaran data dalam bentuk *3D Plot* berdasarkan data *Age* (Umur), variabel *Contact Other Covid* (Kontak dengan pengidap Covid), dan variabel *Covid Result* (Hasil Covid) dengan masing-masing *cluster*. Adapun visualisasinya adalah sebagai berikut :



Gambar 4. Visualisasi 3D Plot

Gambar 5 menunjukkan bahwa posisi *cluster* pada masing-masing 566.602 data kasus berdasarkan warna yang berbeda-beda pula, bulatan kecil adalah posisi data, semetara segitiga adalah titik pusat *cluster*. Pada bulatan kecil dengan warna biru merupakan data-data kasus Covid-19 yang berada pada *cluster-0* yang mana data-data ini memiliki titik pusat *cluster* pada atribut "Age", "Contact Other Covid" dan "Covid Result" yang telah diskalakan secara berurut yaitu, 0.332201, 0.243398, dan 0.376881 dengan total anggota *cluster* sebanyak 214597 data kasus. Untuk *cluster-1* dengan warna merah secara berurut yaitu, 0.331019, 0.269314, dan 0.397774 dengan total anggota *cluster* sebanyak 230216 data kasus. Lalu untuk *cluster-2* dengan warna hijau secara berurut yaitu, 0.441378, 0.513381, dan 0.278937 dengan total anggota *cluster* sebanyak 121789 data kasus.

D. Simpulan

Penerapan algoritma *K-Means* dalam mengklasterisasikan data kasus Covid-19 yang didapat dari kaggle.com berupa data pasien sejak 01 Januari 2020 sampai dengan 31 Mei 2020 sebanyak 566.602 kasus. Hasil *clustering* data kasus yang terkena Covid-19 dengan $k=3$ menghasilkan nilai WCSS sebesar 261181.5. Perhitungan Algoritma *K-Means* dengan bantuan *Google Colaboratory* menghasilkan *cluster* dengan jumlah pasien terkena Covid-19 pada *cluster-0* sebesar 214.597 kasus, pada *cluster-1* sebesar 230.216 kasus, dan pada *cluster-2* sebesar 121.789 kasus. Adapun saran pada penelitian kedepannya diharapkan dapat memanfaatkan *data mining* dan teknik evaluasi yang lain, agar kedepannya dapat dijadikan bahan perbandingan pada hasil penelitian berdasarkan teknik *data mining* dan evaluasi mana yang lebih baik.

E. Ucapan Terima Kasih

Ucapan terima kasih penulis berikan ke berbagai pihak terkait khususnya kepada Tim Penelitian ini, sehingga penelitian ini dapat terselesaikan. Semoga penelitian yang dilakukan ini memiliki banyak kegunaan dan manfaat dalam pengembangan penelitian kedepannya.

F. Referensi

- [1] Susilo, A., Rumende, C. M., Pitoyo, C. W., Santoso, W. D., Yulianti, M., Herikurniawan, Sinto, R., Singh, G., Nainggolan, L., Nelwan, E. J., Chen, L. K., Widhani, A., Wijaya, E., Wicaksana, B., Maksum, M., Annisa, F., Jasirwan, C. O. M., & Yuniastuti, E. "Coronavirus Disease 2019: Tinjauan Literatur Terkini". *Jurnal Penyakit Dalam Indonesia*, 7(1), 45, 2020.
- [2] Fadhilah, M. U., Fauziyah, U., Cahyani, A. A., & Arif, L. "Evaluasi Pelayanan Vaksin Covid-19 (Studi Kasus Puskesmas Mojo Kota Surabaya)". *Journal Publicuho*, 4(2), 536-552, 2021.
- [3] Siburian, T., Safii, M., & Parlina, I. "Penerapan Algoritma K-Means Clustering untuk Pengelompokan Harga Eceran Beras di Pasar Tradisional Berdasarkan Wilayah Kota". *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 927-936, 2019.
- [4] Nawangsih, I., Puspita, R., & Suherman. "Implementasi Algoritma K-Means Dalam Mengkategorikan Produk Terlaris Dan Kurang Laris Pada Toko Alfamart Cikarang". *PELITA TEKNOLOGI*, 16(1), 79-87, 2021.
- [5] Syafitri, S. A. A., Dasriani, N. G. A., & Mayadi. "Pengelompokan Data Pelanggan Indihome Menggunakan Metode K-Means Berbasis Web". *JURNAL UNIVERSITAS BUMIGORA*, 2021.
- [6] Winarta, A., & Kurniawan, W. J. "Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python". *Jurnal Teknik Informatika Kaputama (JTik)*, 5(1), 113-119, 2021.
- [7] Sari, Y. P., Primajaya, A., & Irawan, A. S. Y. "Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang". *INOVTEK Polbeng - Seri Informatika*, 5(2), 229, 2020.
- [8] Siagian, R., Sirait, P., Halim, A., Komputer, P., Area, U. M., & Estate, M. "Penerapan Algoritma K-Means dan K-Medoids untuk Segmentasi Pelanggan pada Data Transaksi E-Commerce". *SISTEMASI: Jurnal Sistem Informasi*, 11, 260-270, 2022.
- [9] Merliana, N. P. E., Ernawati, & Santoso, A. J. "Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering". *Prosiding Seminar Nasional Multi Disiplin Ilmu & Call For Papers UNISBANK (SENDI_U)*, 2018.
- [10] Indriyani, F., & Irfiani, E. "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means". *JUITA: Jurnal Informatika*, 7(2), 109, 2019.
- [11] Rahma, L., Syaputra, H., Mirza, A. H., & Purnamasari, S. D. "Objek Deteksi Makanan Khas Palembang Menggunakan Algoritma YOLO (You Only Look Once)". *Jurnal Nasional Ilmu Komputer*, 2(3), 2021.
- [12] Yunita, F. "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru". *SISTEMASI: Jurnal Sistem Informasi*, 7(3), 238, 2018.

-
- [13] Sulistyawati, A. A. D., & Sadikin, M. "Penerapan Algoritma K-Medoids Untuk Menentukan Segmentasi Pelanggan". *SISTEMASI: Jurnal Sistem Informasi*, 10(3), 2021.
 - [14] Azzahra, A., & Wijayanto, A. W. "Perbandingan Agglomerative Hierarchical dan K-Means dalam Pengelompokan Provinsi Berdasarkan Pelayanan Kesehatan Maternal". *SISTEMASI: Jurnal Sistem Informasi*, 11(2), 481–495, 2022.
 - [15] Ikhsan, E. "Penerapan K-Means Clustering dari Log Data Moodle untuk Menentukan Perilaku Peserta pada Pembelajaran Daring". *SISTEMASI: Jurnal Sistem Informasi*, 10(2), 414, 2021.