

Klusterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means

Achmad Solichin^{1)*}, Khansa Khairunnisa²⁾

^{1), 2)} Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta
achmad.solichin@budiluhur.ac.id^{1)*}, unnicha23@gmail.com²⁾

Abstrak

Corona virus (COVID-19) merupakan jenis virus baru yang ditemukan pada manusia di propinsi Wuhan, Cina pada bulan Desember 2019. Virus ini dapat menular dari manusia ke manusia melalui tetesan kecil (droplet) dari hidung atau mulut pada saat batuk, bersin, atau berbicara. Oleh karena itu, di masa pandemi ini sangat penting untuk menjaga jarak dengan orang lain dan menghindari wilayah dengan persebaran COVID-19 yang tinggi. Pada penelitian ini dilakukan klusterisasi persebaran virus Corona di DKI Jakarta dengan menerapkan metode data mining. Pengelompokan dilakukan berdasarkan parameter jumlah ODP, PDP, kasus Positif, pasien sembuh dan pasien meninggal. Pada penelitian ini, untuk melakukan klusterisasi data digunakan metode K-Means dan metode pengukuran jarak Euclidean. Penelitian ini menghasilkan prototipe aplikasi pengelompokan data persebaran pasien Covid-19. Berdasarkan pengujian, jumlah kluster yang direkomendasikan adalah 9 kluster. Hasil penelitian ini diharapkan dapat membantu pemerintah DKI Jakarta dalam mengambil keputusan strategis dalam mengurangi persebaran virus Corona di DKI Jakarta.

Kata kunci: corona, Jakarta, klusterisasi, k-means

Abstract

[Corona Virus (Covid-19) Clustering in Jakarta using K-Means Method] Corona virus (COVID-19) is a new type of virus found in humans in the province of Wuhan, China in December 2019. This virus can be transmitted from person to person through small droplets from the nose or mouth when coughing, sneezing, or talking. Therefore, during this pandemic, it is very important to keep your distance from other people and avoid areas with a high spread of COVID-19. In this study, the distribution of the Coronavirus in DKI Jakarta was clustered by applying the data mining method. The clustering was carried out based on the parameters of the number of ODP, PDP, positive cases, patients recovered and patients died. In this study, to perform data clustering, the K-Means method and the Euclidean distance measurement method were used. This study produced a prototype application for the distribution of Covid-19 patient distribution data. Based on the test, the recommended number of clusters is 9 clusters. The results of this study are expected to help the DKI Jakarta government in making strategic decisions in reducing the spread of the Corona virus in DKI Jakarta.

Keywords: corona, Jakarta, clustering, k-means

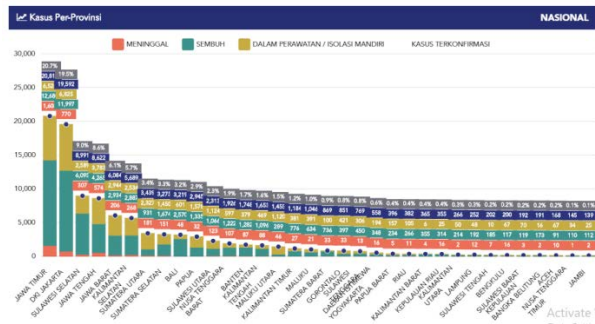
1. PENDAHULUAN

Virus Corona (Covid-19) merupakan salah satu virus mematikan yang dapat menjangkiti manusia dan hewan. Virus tersebut pertama kali ditemukan di propinsi Wuhan, Cina pada akhir Desember 2019. Saat ini dunia sedang mengalami pandemi akibat virus corona tersebut. Seluruh negara sedang berjuang mencegah, menghadapi, dan mengatasi persebaran virus corona. Berdasarkan data WHO, hingga 16 September 2020 terdapat total 29.444.198 kasus terkonfirmasi dari seluruh dunia dengan rata-rata tingkat kematian sebesar 3% [1]. Pandemi Covid-19 juga berdampak ke perekonomian dunia.

Virus corona dapat menular dari manusia ke manusia melalui droplet (tetesan super kecil) yang dihasilkan oleh penderita saat batuk, bersin, atau

berbicara. Beberapa penelitian juga membuktikan bahwa virus corona dapat bertahan beberapa jam di benda-benda yang disentuh oleh penderita. Jika terdapat seseorang menyentuh benda yang sudah terkontaminasi dengan droplet dan menyentuh mata, hidung atau mulut, maka orang itu dapat tertular COVID-19. Oleh karena itu, salah satu pencegahan penularan virus ini adalah dengan menjaga jarak fisik, menggunakan masker dan senantiasa menjaga kesehatan fisik dan non-fisik. Walaupun sejumlah penelitian telah dilakukan untuk mendeteksi virus Covid-19 [2], namun hingga saat ini belum ditemukan cara menghentikan penyebaran virus ini. Pencegahan dapat dilakukan dengan membatasi pergerakan manusia, menjaga jarak fisik, menggunakan masker, dan menerapkan pola hidup sehat.

Provinsi DKI Jakarta merupakan titik awal persebaran virus Corona di Indonesia, dan hingga kini termasuk dalam penyumbang terbesar kasus COVID-19 di Indonesia. Berdasarkan data dari Gugus Tugas Percepatan Penanganan COVID-19[3], per 28 Juli 2020 DKI Jakarta berada di urutan kedua tertinggi dalam total konfirmasi kasus COVID-19 di Indonesia, setelah provinsi Jawa Tengah, dengan total kasus terkonfirmasi COVID-19 sebanyak 19.592 kasus, pasien dalam perawatan/isolasi mandiri sebanyak 6.825 orang, pasien sembuh sebanyak 11.977 orang dan pasien meninggal sebanyak 770 orang. Gambar 1 menampilkan jumlah kasus Covid-19 di Indonesia hingga tanggal 28 Juli 2020.



Gambar 1. Jumlah Kasus Covid-19 di Indonesia [3]

Beberapa penelitian telah melakukan klusterisasi kasus Covid-19 di tingkat dunia [4] dan beberapa negara [5], namun diperlukan analisis lebih mendalam untuk kasus penyebaran Covid-19 di Indonesia [6], khususnya di DKI Jakarta. Tujuan dilakukannya penelitian ini untuk mengkluster persebaran Virus Corona di provinsi DKI Jakarta berdasarkan parameter jumlah ODP, PDP, kasus Positif, pasien sembuh dan pasien meninggal. Klusterisasi dilakukan menggunakan metode K-means, yang mengelompokkan data ke dalam beberapa klaster berdasarkan kemiripan data.

Dengan data persebaran virus Corona yang sudah dikelompokkan menggunakan Algoritma K-means Clustering ini diharapkan dapat membantu pengambilan keputusan yang tepat dalam mengurangi persebaran virus Corona dan meminimalisir jumlah pasien positif COVID-19. Berdasarkan latar belakang yang telah dijabarkan diatas, dapat dirumuskan masalah yang dibahas pada penelitian ini, yaitu bagaimana melakukan klusterisasi persebaran virus Corona di Jakarta dengan menggunakan metode K-means Clustering.

Algoritma klusterisasi K-Means merupakan salah satu dari algoritma data mining yang populer di kalangan peneliti [7]. Jika dibandingkan dengan algoritma klusterisasi lainnya, k-means juga masih cukup tangguh di berbagai jenis data [8].

Beberapa penelitian memanfaatkan algoritma klusterisasi k-means untuk berbagai keperluan, antara lain untuk memprediksi waktu kelulusan mahasiswa [9], estimasi biaya rawat inap rumah sakit [10], pengelompokan penjualan produk [11], pengelompokan karakteristik pengguna sosial media

[12], pengelompokan kunjungan wisatawan [13], dan lain-lain. Tabel x menyajikan rangkuman beberapa penelitian terkini yang memanfaatkan algoritma klusterisasi K-Means.

Tabel 1. Penelitian Terkait Klusterisasi Data Menggunakan Algoritma K-Means

Paper	Tujuan Penelitian	Hasil Penelitian
[9]	Prediksi waktu kelulusan mahasiswa	Pada penelitian ini algoritma k-means berhasil diterapkan untuk memprediksikan kelulusan mahasiswa, namun penelitian tidak mengukur nilai k yang paling optimal.
[10]	Klusterisasi data pasien berdasarkan kemiripan penyakit pasien. Kemiripan diukur menggunakan K-Means dan Jaccard Similarity	Penelitian ini membandingkan metode pengukuran similarity Semantic Similarity dan Jaccard Similarity. Hasilnya metode Semantic Similarity menghasilkan akurasi yang lebih baik dengan akurasi 91%.
[11]	Pengelompokan penjualan produk pada minimarket MM Tika di Kota Bengkulu untuk membantu pemilik toko menentukan strategi penjualan.	Klusterisasi produk dengan menggunakan metode K-Means berdasarkan parameter laku dan tidak laku dapat berfungsi sesuai yang diharapkan.
[12]	Klusterisasi karakteristik pengguna Instagram berdasarkan kecocokan hashtag yang digunakan	Penelitian ini berhasil melakukan klusterisasi pengguna Instagram berdasarkan kecocokan hashtag. Penelitian juga menerapkan metode TF-IDF untuk menyeleksi fitur yang cocok pada proses klusterisasi k-means.
[13]	Analisis penerapan data mining dalam mengelompokkan jumlah kunjungan wisatawan asing ke Prov. DKI Jakarta.	Algoritma k-means berhasil mengelompokkan obyek wisata unggulan di DKI Jakarta menjadi 3 (tiga) klaster.

2. METODE

2.1. Dataset

Penelitian Klasterisasi Persebaran Virus Corona di DKI Jakarta dilakukan menggunakan dataset ODP, PDP dan Positif Covid-19 DKI Jakarta Per-Kecamatan Tahun 2020 dan dataset Rekap Harian Kasus COVID-19 per Kelurahan di Provinsi DKI Jakarta. Data ini diperoleh dari situs Jakarta Open Data [14] yang dikelola oleh Pemerintah Provinsi DKI Jakarta.

Dataset ODP, PDP dan Positif Covid-19 DKI Jakarta Per-Kecamatan Tahun 2020 terdiri dari 4 file data yang dibedakan berdasarkan waktu tercatatnya kasus per bulan, yakni pada bulan Maret, April, Mei dan Juni. Dataset ini terdiri dari beberapa *field* yang dapat dilihat pada Tabel 2.

Tabel 2. Struktur Dataset Persebaran Kasus Covid-19 DKI Jakarta per Kecamatan

Nama Field	Tipe	Keterangan
tanggal	varchar	Tanggal
kecamatan	varchar	Nama kecamatan
odp	int	ODP (Orang Dalam Pemantauan)
proses_pemantauan	int	Proses pemantauan
selesai_pemantauan	int	Selesai pemantauan
pdp	int	PDP (Pasien Dalam Pengawasan)
masih_dirawat	int	Jumlah pasien masih dirawat
pulang_dan_sehat	int	Jumlah pasien yang pulang dan sehat
positif	int	Jumlah yang positif
dirawat	int	Jumlah yang dirawat
sembuh	int	Jumlah yang sehat
meninggal	int	Jumlah yang meninggal
isolasi_mandiri	int	Jumlah yang melakukan isolasi mandiri

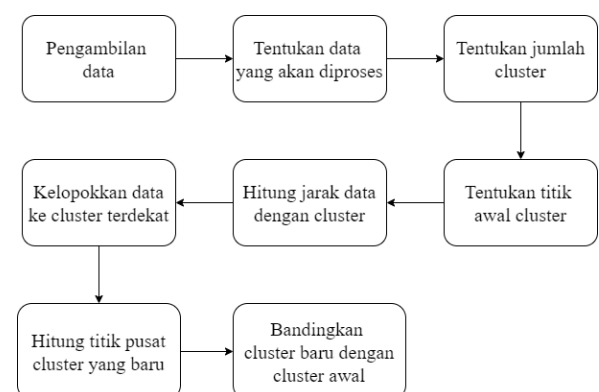
Dataset Rekap Harian Kasus COVID-19 per Kelurahan di Provinsi DKI Jakarta terdiri dari 4 dataset yang dibedakan berdasarkan waktu tercatatnya kasus per bulan, yakni pada bulan Maret, April, Mei dan Juni. Masing-masing dataset ini terbagi lagi berdasarkan waktu tercatatnya kasus per hari. Dataset ini terdiri dari beberapa *field* yang dapat dilihat pada Tabel 3.

Tabel 3. Struktur Dataset Persebaran Kasus Covid-19 DKI Jakarta per Kelurahan

Nama Field	Tipe	Keterangan
id_kel	int	ID Data
nama_provinsi	varchar	Nama Provinsi
nama_kota	varchar	Nama Kota
nama_kecamatan	varchar	Nama Kecamatan
nama_kelurahan	varchar	Nama Kelurahan
odp	int	ODP (Orang Dalam Pemantauan)
proses_pemantauan	int	Proses pemantauan
selesai_pemantauan	int	Selesai pemantauan
pdp	int	PDP (Pasien Dalam Pengawasan)
masih_dirawat	int	Jumlah pasien masih dirawat
pulang_dan_sehat	int	Jumlah pasien yang pulang dan sehat
positif	int	Jumlah yang positif
dirawat	int	Jumlah yang dirawat
sembuh	int	Jumlah yang sehat
meninggal	int	Jumlah yang meninggal
self_isolation	int	Jumlah yang melakukan isolasi mandiri
keterangan	int	Keterangan

2.2. Penerapan Metode

Penerapan metode K-means pada penelitian Klasterisasi Persebaran Virus Corona di DKI Jakarta dilakukan dalam beberapa tahap seperti disajikan pada Gambar 2.



Gambar 2. Implementasi Metode K-Means

Berikut penjelasan tiap tahapan penerapan metode k-means yang dilakukan pada penelitian ini.

1) Pengambilan Data

Data penelitian ini diperoleh dari situs Jakarta Opendata [14]. Data yang diperoleh disesuaikan dengan spesifikasi basis data. Data tersebut kemudian disimpan ke dalam pangkalan data yang telah dibuat.

2) Seleksi Dataset

User dapat menentukan data yang ingin diproses berdasarkan tanggal data. Dataset yang digunakan bersifat kumulatif, sehingga data yang dipilih adalah total seluruh kasus sampai dengan tanggal yang dipilih.

3) Menentukan Jumlah Kluster

User dapat menentukan berapa jumlah kluster yang ingin dibentuk pada tiap proses. Pada penelitian dibatasi jumlah kluster minimal 2 dan jumlah maksimal kluster adalah 10.

4) Menentukan Titik Awal Kluster (Centroid)

Titik pusat kluster awal, atau dapat juga disebut centroid awal, ditentukan secara random berdasarkan jumlah kluster dan jumlah data yang akan diproses.

5) Menghitung Jarak Setiap Data dengan Titik Pusat Kluster

Jarak antar setiap data dengan masing-masing kluster dihitung menggunakan rumus Euclidean Distance (D) seperti disajikan pada Persamaan (1).

$$D_{(i,j)} = \sqrt{\sum_{k=1}^n (X_{ik} - C_{jk})^2} \quad (1)$$

yang mana

D = jarak kluster

X_{ik} = nilai data (i,k)

C_{jk} = nilai centroid (j,k)

n = jumlah kluster

6) Pengelompokan Data Berdasarkan Kluster Terdekat

Perhatikan kluster mana yang memiliki jarak terdekat dengan data. Kemudian kelompokkan data ke dalam kluster tersebut.

7) Menghitung Titik Pusat Kluster yang Baru

Setelah semua data dikelompokkan ke dalam kluster, hitung titik pusat kluster yang baru dengan menghitung rata-rata jarak data dengan titik pusat kluster menggunakan Persamaan (2).

$$C_i = \frac{\sum d_i}{n_k} \quad (2)$$

8) Bandingkan Kluster Baru dengan Kluster Awal

Jika kluster baru yang dibentuk memiliki centroid yang berbeda dengan kluster awal, ulangi kembali proses mulai dari tahap 5. Jika centroid kluster baru sudah sama dengan kluster sebelumnya, proses dapat dihentikan dan diperoleh hasil akhir clustering.

2.3. Rancangan Pengujian

Pada penelitian ini, dilakukan pengujian terhadap aplikasi pengelompokan persebaran virus Corona menggunakan Metode *Elbow*. Metode *Elbow* merupakan salah satu metode yang dapat digunakan untuk menguji efektifitas jumlah kluster yang digunakan pada proses klusterisasi. Metode ini bekerja dengan cara melihat persentase hasil perbandingan antara jumlah kluster yang akan membentuk siku pada suatu titik. Jika nilai kluster pertama dengan nilai kluster kedua nilainya mengalami penurunan paling besar maka nilai kluster tersebut yang terbaik.

Untuk mendapatkan perbandingannya adalah dengan menghitung SSE (*Sum of Square Error*) dari masing-masing nilai kluster menggunakan Persamaan (3). Semakin besar jumlah kluster (k) maka nilai SSE akan semakin kecil. Persamaan SSE pada K-Means disajikan pada Persamaan (3).

$$SSE = \sum_{K=1}^K \sum_{x_i \in S_k} \|X_i - C_k\|^2 \quad (3)$$

yang mana:

K = jumlah kluster

x_i = data ke i

C_k = centroid kluster

Setelah dilakukan perhitungan nilai SSE, nilai kluster k dapat mengalami perubahan (penurunan). Nilai K akan turun secara bertahap hingga hasil dari nilai K tersebut memiliki nilai yang stabil. Misalnya nilai kluster K=2 ke K=3, kemudian dari K=3 ke K=4, terlihat penurunan drastis membentuk siku pada titik K=3 maka nilai kluster k yang ideal adalah K=3.

Penerapan metode *Elbow* dilakukan dalam beberapa tahap :

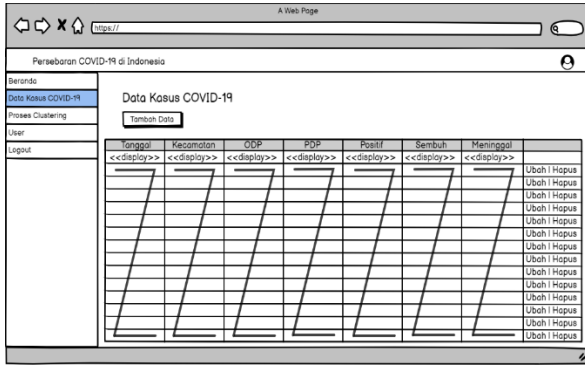
- 1) Tentukan range nilai K yang akan diuji.
- 2) Hitung hasil *Sum of Square Error* dari tiap nilai K.
- 3) Hitung selisih hasil *Sum of Square Error* dari tiap nilai K.
- 4) Tetapkan nilai K yang memiliki selisih terbesar sebagai nilai K terbaik.

2.4. Rancangan Prototipe

Pada penelitian ini dikembangkan sebuah prototype aplikasi pengklusteran data Covid-19 menggunakan metode K-Means. Prototipe dikembangkan menggunakan bahasa pemrograman PHP dan perangkat pangkalan data MySQL. Keduanya

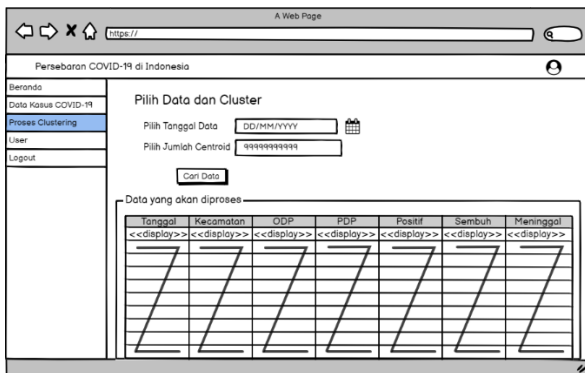
dipilih atas dasar kemudahan dan rendahnya biaya dalam implementasi [15].

Pada halaman data kasus Covid-19, ditampilkan data kasus yang sudah diinput sebelumnya. Data kasus ditampilkan dalam bentuk tabel. User juga dapat menambah data kasus baru dengan menekan tombol "Tambah Data". Gambar 3 menyajikan rancangan layar data kasus Covid-19 yang disajikan untuk setiap kecamatan di DKI Jakarta.



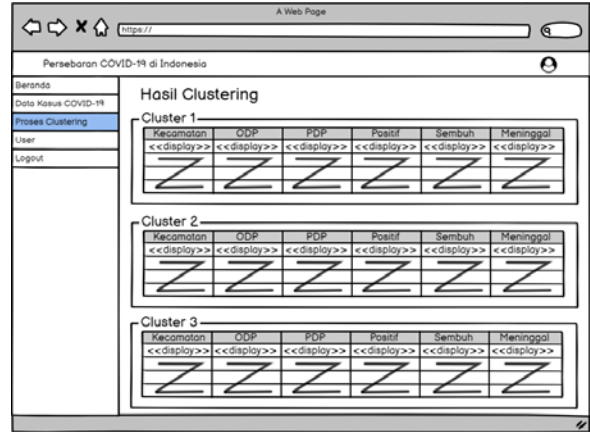
Gambar 3. Rancangan Layar Data Kasus COVID-19

Sebelum melakukan proses clustering, pengguna harus memilih tanggal data dan jumlah kluster terlebih dahulu. Setelah menekan tombol "Cari Data", akan ditampilkan data yang akan diproses dalam bentuk tabel. Gambar 4 menyajikan rancangan layar pemilihan tanggal data dan jumlah kluster atau centroid.



Gambar 4. Rancangan Layar Pemilihan Dataset

Setelah proses klusterisasi dilakukan, pada Gambar 5 ditampilkan hasil akhir proses clustering dalam bentuk tabel sesuai dengan jumlah klaster yang dipilih.

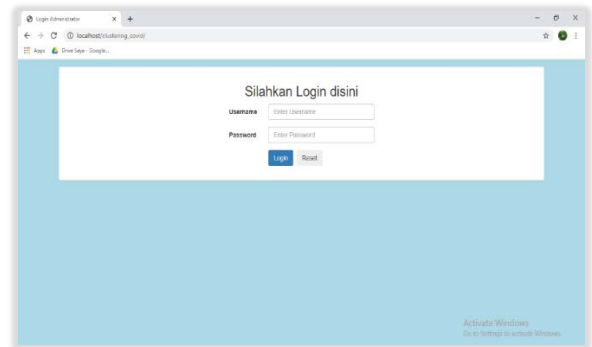


Gambar 5. Rancangan Layar Hasil Klusterisasi

3. HASIL DAN PEMBAHASAN

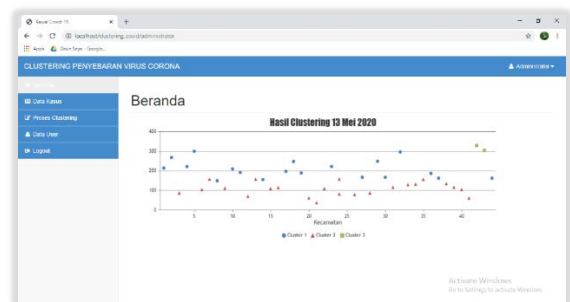
3.1. Prototipe Aplikasi

Berdasarkan rancangan layar dan rancangan aplikasi yang sudah dijelaskan pada bagian sebelumnya, dikembangkan prototipe aplikasi. Gambar 6 merupakan tampilan layar halaman login. Saat memulai aplikasi, pengguna diarahkan ke halaman Login. Pada halaman ini, pengguna harus menginput username dan password sebelum menjalankan aplikasi.



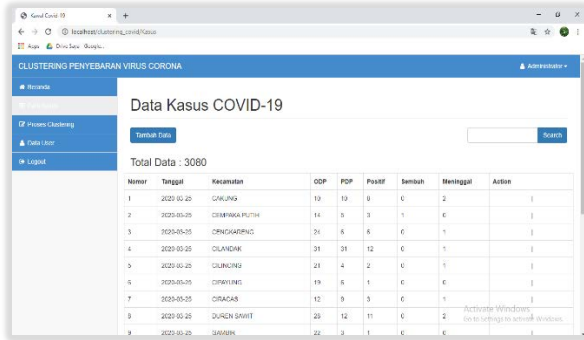
Gambar 6. Tampilan Layar Login ke Aplikasi

Setelah berhasil login, pengguna diarahkan ke halaman Beranda seperti disajikan pada Gambar 7. Pada halaman ini ditampilkan hasil proses clustering yang terakhir dilakukan.



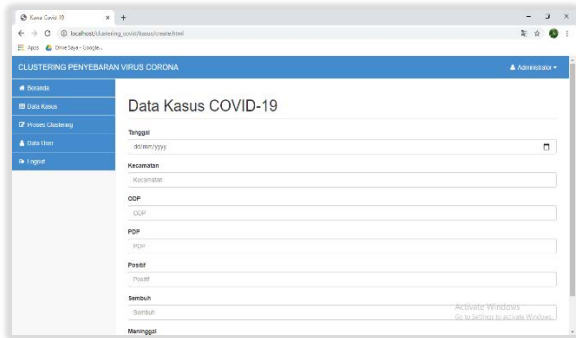
Gambar 7. Tampilan Layar Dashboard

Selanjutnya, pada halaman data kasus COVID-19 seperti pada Gambar 8, ditampilkan data kasus yang sudah diinput sebelumnya. Data kasus ditampilkan dalam bentuk tabel. Pengguna juga dapat menambah data kasus baru dengan menekan tombol “Tambah Data”. Selain itu, Pengguna dapat mencari data berdasarkan kecamatan atau tanggal kasus.



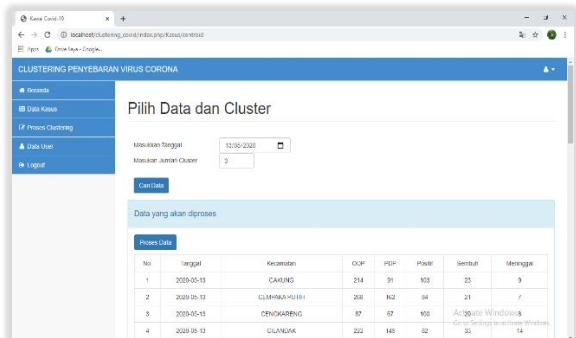
Gambar 8. Tampilan Layar Data Kasus COVID-19

Untuk menambah data kasus, pengguna harus mengisi semua kolom yang tersedia, yakni kolom Tanggal Kasus, Kecamatan, ODP, PDP, Positif, Sembuh dan Meninggal. Gambar 9 merupakan tampilan halaman pengisian kasus baru.



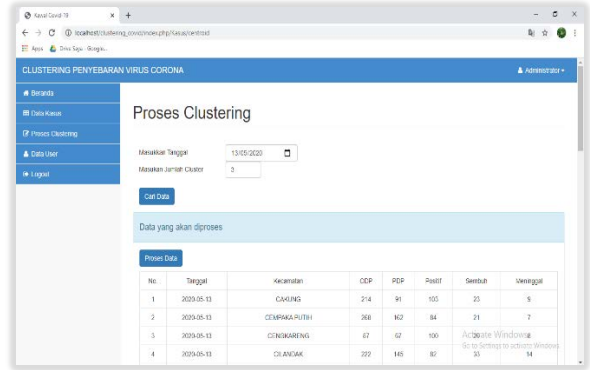
Gambar 9. Tampilan Layar Tambah Data Kasus

Sebelum melakukan proses clustering pengguna harus memilih tanggal data dan jumlah kluster terlebih dahulu. Setelah menekan tombol “Cari Data”, akan ditampilkan data yang akan diproses dalam bentuk tabel. Untuk memulai proses clustering, tekan tombol “Proses Data”. Gambar 10 menampilkan tampilan layar pemilihan dataset berdasarkan tanggal kasus.



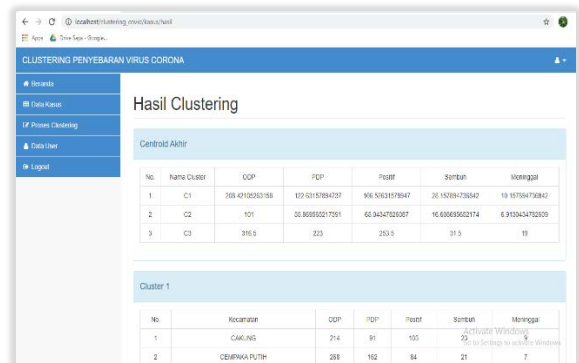
Gambar 10. Tampilan Layar Pemilihan Dataset

Gambar 11 menyajikan halaman Proses Clustering yang ditampilkan nilai centroid dan hasil dari tiap iterasi dalam bentuk tabel. Pada halaman ini juga terdapat tombol “Tampilkan Hasil Clustering” untuk menampilkan hasil akhir dari proses clustering yang telah dilakukan.



Gambar 11. Tampilan Layar Proses Clustering

Selanjutnya, Gambar 12 ditampilkan hasil akhir proses clustering dalam bentuk tabel. Hasil klusterisasi menyajikan pengelompokan wilayah di DKI Jakarta berdasarkan jumlah kluster yang ditentukan pada saat memulai proses pengklusteran.



Gambar 12. Tampilan Layar Hasil Klusterisasi

3.2. Pengujian Kluster Terbaik

Pengujian yang dilakukan pada penelitian ini bertujuan untuk mencari jumlah kluster terbaik yang bisa dibentuk dari data kasus COVID-19 di DKI Jakarta. Berikut tahapan pengujianya.

- 1) **Tentukan range nilai K yang akan diuji**
 Pada pengujian ini, jumlah kluster yang akan diuji adalah dari $K = \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Pengujian menggunakan data uji kasus Covid-19 per tanggal 13 Mei 2020.
- 2) **Hitung hasil dan selisih Sum of Square Error (SSE) dari tiap nilai K**
 Perhitungan nilai SSE dan selisihnya untuk setiap jumlah kluster dilakukan menggunakan Persamaan (3). Berdasarkan hasil pengujian diperoleh hasil perhitungan nilai SSE dan selisih untuk setiap nilai K seperti disajikan pada Tabel 4.

Tabel 4. Hasil Perhitungan *Sum of Square Error* (SSE)

Nilai K	Besar SSE	Selisih
2	227.714,870	0
3	149.099,212	78.615,658
4	155.351,429	6.252,217
5	93.951,443	61.399,985
6	105.038,753	11.087,310
7	125.885,000	20.846,247
8	567.909,010	442.024,010
9	87.142,675	480.766,335
10	75.412,235	11.730,441

3) **Tetapkan nilai K yang memiliki selisih terbesar sebagai nilai K terbaik.**

Berdasarkan hasil pengujian pada Tabel 4, dapat disimpulkan bahwa jumlah kluster yang memiliki nilai selisih terbesar adalah 9 kluster. Oleh karena itu, jumlah kluster tersebut direkomendasikan untuk digunakan dalam pengelompokan kasus Covid-19 di DKI Jakarta. Namun demikian, pembagian kluster tersebut tentunya harus memperhatikan pertimbangan lainnya. Jika pembagian kluster terlalu banyak, dapat menimbulkan kebingungan bagi warga DKI Jakarta.

3.3. Pembahasan Hasil Penelitian

Penelitian ini memiliki kontribusi praktis dalam membantu penyajian data kasus Covid-19 secara lebih baik. Prototipe yang dikembangkan dapat menyajikan hasil klusterisasi data kasus Covid-19 di Jakarta yang ditampilkan per kecamatan. Pengguna aplikasi dapat menentukan jumlah kluster dan data yang akan dikluster. Pada penelitian ini dilakukan pengujian jumlah kluster terbaik menggunakan metode Elbow. Hasil pengujian menunjukkan bahwa jumlah kluster yang direkomendasikan adalah 9 (sembilan) kluster (lihat Tabel 4). Jumlah tersebut dipandang terlalu besar karena tidak efektif dan kurang baik untuk keperluan komunikasi publik.

Jika melihat data pada Tabel 4, pengelompokan data menjadi 3 (tiga) kluster juga dapat dipilih karena memiliki selisih SSE yang cukup besar dibanding yang lainnya. Dengan 3 (tiga) kluster, penyajian data menjadi lebih efektif dan informatif. Tabel 5 menyajikan hasil klusterisasi data jika dilakukan dengan nilai $k=3$ atau dengan 3 kluster.

Klusterisasi data kasus Covid-19 sudah dilakukan oleh beberapa peneliti. Namun demikian, di tingkat propinsi DKI Jakarta belum ada yang melakukannya. Pada penelitian oleh Gunawan dkk mengusulkan algoritma klusterisasi k-medoids untuk data seluruh Indonesia yang disajikan per propinsi [16]. Penelitian ini berusaha menyelesaikan permasalahan klusterisasi di tingkat wilayah Propinsi DKI Jakarta.

Tabel 5. Klusterisasi data kasus dengan nilai $k = 3$

Kluster	Kecamatan
C1	Cakung, Cempaka Putih, Cilandak, Cilincing, Duren Sawit, Grogol, Petamburan, Jagakarsa, Kali Deres, Kebon Jeruk, Kelapa Gading, Kemayoran, Koja, Matraman, Pademangan, Palmerah, Pasar Minggu, Pulo Gadung, Sawah Besar, Tebet
C2	Cengkareng, Cipayung, Ciracas, Gambir, Jatinegara, Johar Baru, Kebayoran Baru, Kebayoran Lama, Kembangan, Kep. Seribu Selatan, Kep. Seribu Utara, Kramat Jati, Makasar, Mampang Prapatan, Menteng, Pancoran, Pasar Rebo, Penjarangan, Pesanggrahan, Senen, Setia Budi, Taman Sari, Tambora
C3	Tanah Abang, Tanjung Priok

4. KESIMPULAN

Berdasarkan hasil pengujian dan uji coba terhadap penerapan metode k-means untuk melakukan pengelompokan atau klusterisasi kasus Covid-19 di DKI Jakarta, dapat disimpulkan bahwa metode k-means dapat digunakan dengan baik. Adapun jumlah kluster yang direkomendasikan berdasarkan perhitungan nilai SSE adalah 9 kluster. Namun demikian pada prakteknya, jumlah kluster tersebut dapat ditentukan sendiri oleh pengguna aplikasi berdasarkan pertimbangan-pertimbangan tertentu. Jika jumlah 9 kluster terlalu banyak, penelitian ini merekomendasikan penggunaan 3 (tiga) kluster karena jumlah tersebut memiliki kualitas yang cukup baik berdasarkan hasil perhitungan metode Elbow. Pada klusterisasi dengan 3 (tiga) kluster, dihasilkan pengelompokan C1 sebanyak 19 kecamatan, C2 sebanyak 23 kecamatan, dan C3 sebanyak 2 kecamatan.

Pada penelitian selanjutnya, klusterisasi dapat dikembangkan dengan menerapkan berbagai metode klusterisasi lainnya seperti Fuzzy K-Means, Fuzzy C-Means, K-Medoids, dan sebagainya.

5. DAFTAR PUSTAKA

- [1] WHO, "WHO Coronavirus Disease (COVID-19) Dashboard," *WHO*, 2020. [Daring]. Tersedia pada: <https://covid19.who.int/>. [Diakses: 17-Sep-2020].
- [2] W. Swastika, "Studi Awal Deteksi Covid-19 Menggunakan Citra CT Berbasis Deep Learning," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, hal. 629–634, 2020.
- [3] Gugus Tugas Percepatan Penanganan COVID-19, "Peta Sebaran Covid-19," *covid19.go.id*, 2020. [Daring]. Tersedia pada: <https://covid19.go.id/peta-sebaran>. [Diakses: 28-Jul-2020].
- [4] V. Zarikas, S. G. Pouloupoulos, Z. Gareiou, dan E. Zervas, "Clustering analysis of countries using the COVID-19 cases dataset," *Data Br.*,

- vol. 31, hal. 2352–3409, 2020.
- [5] M. Azarafza, M. Azarafza, dan H. Akgün, “Clustering method for spread pattern analysis of corona-virus (COVID-19) infection in Iran,” *J. Appl. Sci. Eng. Technol. Educ*, vol. 3, no. 1, hal. 1–6, 2020.
- [6] G. D. Rembulan, T. Wijaya, D. Palullungan, K. N. Alfina, dan M. Qurthuby, “Kebijakan Pemerintah Mengenai Coronavirus Disease (COVID-19) di Setiap Provinsi di Indonesia Berdasarkan Analisis Klaster Government Policy Regarding Coronavirus Disease (COVID-19) in Each Province in Indonesia Based on Clustering Analysis,” *J. Ind. Eng. Manag. Syst.*, vol. 13, no. 2, hal. 74–86, 2020.
- [7] X. Wu *et al.*, “Top 10 algorithms in data mining,” in *Knowledge and Information Systems*, 2008, vol. 14, no. 1, hal. 1–37.
- [8] M. Z. Rodriguez *et al.*, *Clustering algorithms: A comparative approach*, vol. 14, no. 1. 2019.
- [9] H. Priyatman, F. Sajid, dan D. Haldivany, “Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa,” *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 1, hal. 62, 2019.
- [10] I. B. G. Sarasvananda, R. Wardoyo, dan A. K. Sari, “The K-Means Clustering Algorithm With Semantic Similarity To Estimate The Cost of Hospitalization,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 4, hal. 313, 2019.
- [11] Y. Darmi dan A. Setiawan, “Penerapan Metode Clustering K-Means dalam Pengelompokan Penjualan Produk,” *J. Media Infotama*, vol. 12, no. 2, hal. 148–157, 2016.
- [12] M. Habibi dan P. W. Cahyo, “Clustering User Characteristics Based on the influence of Hashtags on the Instagram Platform,” *Indones. J. Comput. Cybern. Syst.*, vol. 13, no. 4, hal. 399–408, 2019.
- [13] L. Maulida, “Penerapan Datamining dalam Mengelompokkan Kunjungan Wisatawan ke Objek Wisata Unggulan di Prov. DKI Jakarta dengan K-Means,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 2, no. 3, hal. 167–174, 2018.
- [14] Jakarta Open Data, “Data Kasus Positif Covid-19 Provinsi DKI Jakarta Tahun 2020,” *data.jakarta.go.id*, 2020. [Daring]. Tersedia pada: <https://data.jakarta.go.id/dataset/data-kasus-positif-covid-19-dki-jakarta>. [Diakses: 28-Jul-2020].
- [15] A. Solichin, *Pemrograman Web dengan PHP dan MySQL*. Budi Luhur Press, 2016.
- [16] I. Gunawan, G. Anggraeni, E. S. Rini, Y. M. Putri, dan Y. K. Zikri, “Klasterisasi provinsi di Indonesia berbasis perkembangan kasus Covid-19 menggunakan metode K-Medoids,” in *Seminar Nasional Matematika dan Pendidikan Matematika (SENATIK)*, 2020, hal. 301–306.