

## DATA MINING APPLICATION FOR CLUSTERING *COVID-19* SPREAD AREAS IN DKI JAKARTA USING THE K-MEANS ALGORITHM

**Turkhamun Adi Kurniawan<sup>1</sup>, T.W. Wisjhnuadji<sup>2</sup>, Habib Kholil Al Hanif<sup>3</sup>**

*Fakultas Teknik Informatika*

*Universitas Satya Negara Indonesia<sup>1</sup>,*

*Fakultas Teknik Informatika Universitas Budi Luhur<sup>2</sup>, Fakultas Teknik Informatika*

*Universitas Satya Negara Indonesia<sup>3</sup>*

Jl. Sultan Iskandar Muda No. 11, Jakarta Selatan 12240

E-mail: <sup>1</sup>t.adikurniawan@usni.ac.id <sup>2</sup>wisjhnuadji@budituluhur.ac.id, <sup>3</sup>yekonur@rocketmail.com

### *Abstract*

Coronavirus Disease 2019 (*coronavirus disease2019*, abbreviated as *COVID-19*) is an infectious disease caused by *SARS-CoV-2*, a type of coronavirus. Covid-19 patients can experience fever, dry batik, and difficulty breathing. The infection spreads from one person to another through a splash (*droplet*) from the respiratory tract produced when coughing or sneezing. The number of residents until 2019 reached 11,063,324 people spread across 6 cities consisting of 44 districts and 267 urban villages, making Covid-19 easy to spread. To be able to see the area of spread of Covid-19, it is necessary to group based on the attributes used consisting of Suspect Cases, *Probable*, Cases, Close Contacts, Confirmed Cases and Deaths. In this study, to cluster the data, the K-Means method and the Euclidean distance measurement method were used. This study produced a prototype application for grouping data on the distribution of Covid-19 patients. The result of the implementation of the K-Means Algorithm is that the Covid-19 spread cluster in DKI Jakarta is divided into 3 (three) clusters, namely cluster 1, cluster 2 and cluster 3. Cluster 1 is a medium case zone, Cluster 2 is a high case zone and Cluster 3 is a low case zone.

**Keywords:** *Covid-19, K-MEANS Algorithm*

### I. PENDAHULUAN

Mesin merupakan salah satu aset perusahaan yang sangat penting untuk menjaga keberlangsungan Penyakit Koronavirus 2019 (*coronavirus disease2019*, disingkat *COVID-19*) merupakan penyakit menular yang disebabkan oleh *SARS-CoV-2*, salah satu jenis korona virus. Penderita Covid-19 bisa mengalami demam, batuk kering, dan kesulitan bernapas. Infeksi menyebar dari satu orang ke orang lainnya melalui percikan (*droplet*) dari saluran pernapasan yang dikeluarkan saat batuk atau bersin. Beberapa penelitian juga membuktikan bahwa virus corona dapat bertahan beberapa jam di benda-benda yang disentuh oleh penderita. Jika seseorang menyentuh benda yang sudah terkontaminasi dengan droplet yang mengandung virus 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. Menurut data dari Badan Pusat Statistik (BPS) banyaknya Jumlah penduduk di DKI Jakarta hingga tahun 2021 mencapai 10,609,681 jiwa yang tersebar di 6 Kota terdiri dari 44 Kecamatan serta 267 Kelurahan. Banyaknya penduduk dan padatnya aktivitas di DKI Jakarta membuat virus covid-19 dapat menular dengan cepat dan tidak terkendali

Provinsi DKI Jakarta merupakan penyumbang terbesar kasus COVID19 di Indonesia. Berdasarkan data dari Gugus Tugas Percepatan Penanganan COVID-19. Beberapa penelitian telah melakukan klasterisasi kasus Covid-19 di tingkat dunia dan beberapa negara, namun diperlukan analisis lebih mendalam untuk kasus penyebaran Covid-19 di Indonesia, khususnya di DKI Jakarta. Sejak diumumkan kasus pertama Covid-19 di DKI Jakarta, data-data yang berkaitan dengan jumlah kasus Covid-19 sudah mulai dikumpulkan (rekap) seperti jumlah kasus positif, PDP, ODP dan meninggal dunia termasuk penyebarannya di seluruh daerah di DKI Jakarta. Data jumlah kasus Covid-19 di DKI Jakarta dapat dimanfaatkan untuk memprediksi jumlah kasus dan membentuk cluster zona terbaru berdasarkan jumlah Kasus Suspek, Probable, Meninggal, Kontak Erat dan Kasus Konfirmasi atau positif dengan menggunakan algoritma clustering K-means. Salah satu algoritma yang termasuk dalam clustering adalah K-means. K-means adalah algoritma yang mengelompokan data dengan memaksimalkan tingkat kemiripan antar cluster. Dengan dibuatnya data persebaran virus Corona yang sudah dikelompokan atau dicluster menggunakan Algoritma ini diharapkan dapat membantu pengambilan keputusan yang tepat dalam mengurangi persebaran virus Covid

## II. LANDASAN TEORI

### A. Data Mining

Penggalian data atau dalam bahasa Inggris disebut data mining merupakan ekstraksi pola yang menarik dari data dengan jumlah besar. Suatu pola dikatakan menarik apabila pola tersebut tidak mudah, tidak diketahui sebelumnya, dan bermanfaat. Pola yang diberikan harus mudah dipahami,.Data mining mempunyai beberapa nama lain, meskipun definisi eksaknya tidak sama, seperti KDD (Knowledge Discovery in Database), analisis pola, arkeologi data, pemanenan info, serta intelegensi bisnis. penggalian data diharapkan ketika data yang tersedia terlalu banyak (contohnya data yang diperoleh dari sistem basis data perusahaan, e-commerce, data saham, data sensus serta data bioinformatika), namun tidak memahami pola apa yang mampu dihasilkan. Data mining sangat perlu dilakukan terutama untuk mengelola data yang sangat besar untuk buat memudahkan kegiatan recording suatu transaksi serta buat proses data warehousing agar bisa menyampaikan info yang akurat bagi penggunanya alasan utama mengapa data mining sangat menarik perhatian industri informasi dalam beberapa tahun belakangan ini ialah karena tersedianya data dalam jumlah yang sangat besar serta semakin besarnya kebutuhan untuk mengganti data tersebut menjadi informasi serta pengetahuan yang bermanfaat karena sesuai penekanan bidang ilmu yaitu melakukan aktivitas mengekstrasi atau menambah pengetahuan dari data yang ukurannya/berjumlah besar , informasi inilah yang nantinya sangat bermanfaat untuk pengembangan. Data mining ialah proses analitik yang didesain untuk memeriksa sejumlah data yang besar dalam mencari suatu pengetahuan tersembunyi yang berharga serta konsisten [Florin G. 2011]. Tujuan dari data mining ialah mencari trend atau pola yang diinginkan pada database besar untuk membantu pada pengambilan keputusan di waktu kedepannya.

### B. Clustering

Clustering adalah suatu proses pengelompokan rekam, observasi, atau mengelompokkan kelas yang mempunyai kemiripan objek [Larose, 2015]. Yang membedakan clustering dengan klasifikasi yaitu tidak adanya variable sasaran dalam melakukan pengelompokan pada proses clustering. Clustering seringkali dilakukan untuk Langkah pertama pada proses data mining. terdapat banyaknya algoritma clustering yang sudah digunakan oleh peneliti sebelumnya seperti Improved K-Means, K-Means, Fuzzy C-Means, DBSCAN,K-means (PAM), CLARANS serta FuzzySubtractive. Dari algoritma yang baru saja disebut mempunyai kelebihan serta kekurangannya masing-masing,tetapi prinsip prosedur pemecahannya sama, yaitu mengelompokkan data sesuai dengan karakteristik serta menghitung jarak kemiripan antar data di satu kelompok.

### C. Algoritma K-means Clustering

Algoritma K-means adalah salah satu algoritma analisis kluster (*cluster analysis*) non hierarki yang bertujuan mendapatkan data dengan memaksimalkan kemiripan karakteristik dalam cluster. Prinsip utama K-Means adalah menyusun k prototype atau pusat massa (centroid) dari sekumpulan data berdimensi. Sebelum diterapkan proses algoritma K-means, dokumen akan di preprocessing terlebih dahulu. Kemudian dokumen direpresentasikan sebagai vektor yang memiliki term dengan nilai tertentu. Algoritma k-means merupakan algoritma yang membutuhkan parameter input sebanyak k dan membagi sekumpulan n objek kedalam k cluster sehingga tingkat kemiripan antar anggota dalam satu cluster tinggi sedangkan tingkat kemiripan dengan anggota pada cluster lain sangat rendah. Kemiripan anggota terhadap cluster diukur dengan kedekatan objek terhadap nilai mean pada cluster atau dapat disebut sebagai centroid cluster

### D. COVID-19

Korona virus 2019 (coronavirus disease 2019, disingkat COVID-19) ialah salah satu jenis korona virus yang merupakan penyakit menular dan disebabkan oleh SARS-CoV-2 [Calisher, Alexander E. (11 Februari 2021)]. Korona virus adalah kumpulan virus yang menginfeksi sistem pernapasan. Pada banyak kasus, virus ini hanya menyebabkan infeksi pernapasan ringan, Seperti flu. Namun, virus ini juga bisa menyebabkan infeksi pernapasan berat, seperti infeksi paru-paru (pneumonia). Penderita Covid-19 bisa mengalami batuk kering,demam serta sulit bernapas. Infeksi menyebar dari satu orang ke orang lain lewat percikan (droplet) melalui saluran pernapasan yang dikeluarkan ketika batuk atau bersin

### E. Kasus suspek

Kasus Suspek Seseorang yang memiliki salah satu dari kriteria berikut:

1. Orang dengan Infeksi Saluran Pernapasan Akut (ISPA) dan pada 14 hari terakhir sebelum timbul gejala memiliki riwayat perjalanan atau tinggal di negara/wilayah Indonesia yang melaporkan transmisi local.

2. Orang dengan salah satu gejala/tanda ISPA, dan pada 14 hari terakhir sebelum timbul gejala memiliki riwayat kontak dengan kasus konfirmasi/probable COVID-19.
3. Orang dengan ISPA berat/pneumonia berat yang membutuhkan perawatan di rumah sakit dan tidak ada penyebab lain berdasarkan gambaran klinis yang meyakinkan

### III. METODE PENELITIAN

#### A. Penerapan Metode

Penerapan metode algoritma K-means dalam penelitian Clustering Persebaran Virus Corona di DKI Jakarta dilakukan beberapa tahap, yaitu:

1. Berikan nilai awal pusat cluster sejumlah  $k$  (banyaknya cluster). pusat cluster diambil dari objek yang dipilih secara acak yang menjadi perwakilan centroid.
2. Tentukan banyaknya setiap titik data (objek) dengan medoid yang paling mirip dengan menggunakan ukuran jarak serta menghitung cost. Berikut persamaan jarak Euclidian Distance:

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

yang mana

$D$  = jarak klaster

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

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

$n$  = jumlah klaster

3. Kelompokkan data ke dalam cluster yang mempunyai jarak yang paling pendek.
4. Hitung pusat cluster.
5. Ulangi Langkah ke 2 - ke 4 hingga sudah tidak ada lagi data yang berpindah ke cluster yang lain

#### B. Kerangka Berpikir

Berdasarkan kerangka berpikir diatas dijelaskan bahwa permasalahan yang menjadi dasar perencanaan ini adalah penyebaran covid-19 di DKI Jakarta. Maka dari itu dibuatkan rumusan masalah yaitu bagaimana cara melakukan clustering terhadap penyebaran virus covid-19 di DKI Jakarta dengan menggunakan algoritma K-Means. Dari rumusan masalah yang ada, diberikan usulan yaitu membuat aplikasi data mining untuk clustering daerah penyebaran covid-19 di DKI Jakarta menggunakan algoritma K-Means. Setelah usulan diberikan, peneliti melakukan teknik pengumpulan data, yaitu pengumpulan data set dan studi pustaka yang bertujuan untuk mendapatkan data yang dibutuhkan untuk clustering covid-19 di DKI Jakarta. Tahap selanjutnya yaitu merancang aplikasi dan pengujian sistem agar aplikasi berjalan dengan baik. peneliti tertarik membuat aplikasi untuk clustering covid-19 untuk bisa melihat penyebaran covid 19 di DKI Jakarta

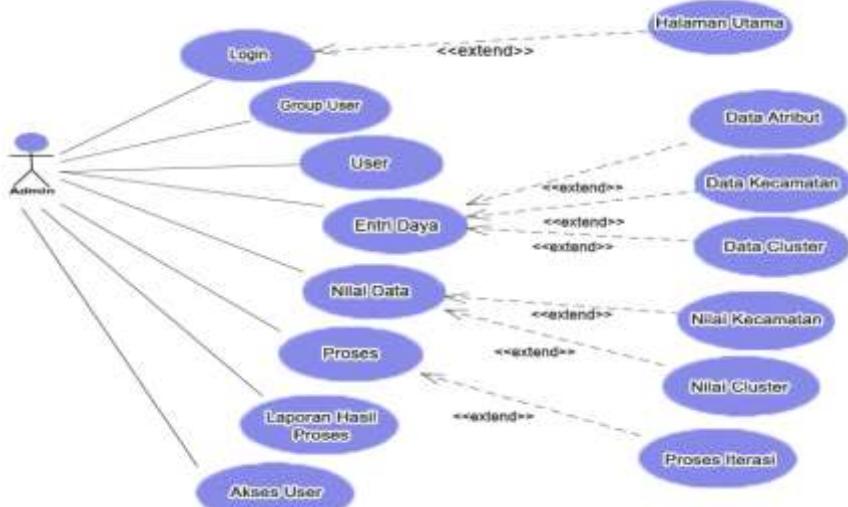


Gambar 1 Sistem yang Diusulkan

#### IV. ANALISA DAN PERANCANGAN SISTEM

##### A. Usecase Diagram.

Berikut adalah Use Case Diagram dari Aplikasi Clustering Covid-19

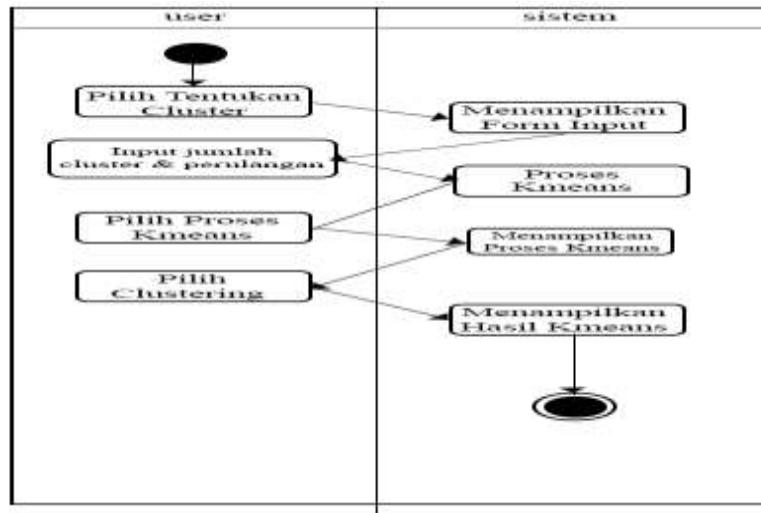


Gambar 2 Use Case Diagram system yang diusulkan

Penjelasan dari use case diagram diatas ialah terdapat 1 aktor yaitu admin, admin dapat melakukan proses login, mengelola user, entri data atau memasukan data, mengolah nilai data, memproses data, menampilkan hasil proses serta akses user. Ditambah juga dengan relasi extend dimana use case dapat berdiri sendiri

##### B. Activity Diagram

Activity diagram merupakan bentuk visual dari alur kerja yang merupakan Tindakan. Berikut Activity diagram dari aplikasi clustering Covid-19.



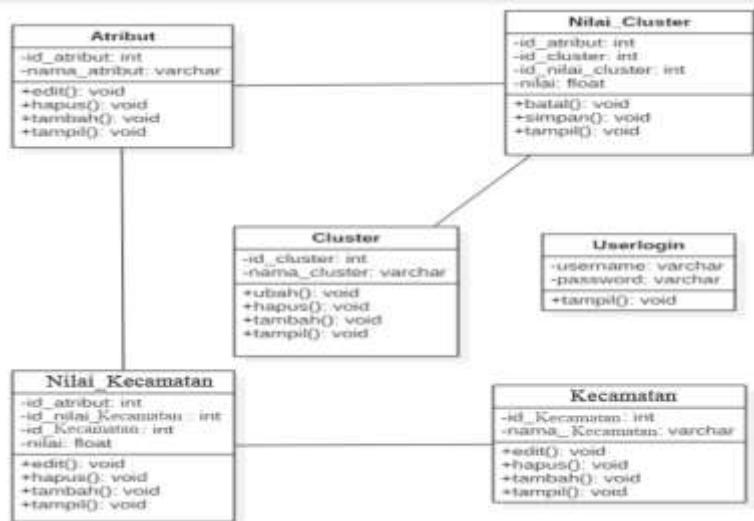
Gambar 3 Activity diagram Pross

Berdasarkan gambar activity diagram pada maintenance terdapat:

1. Satu Initial mode dimana objek memulai kegiatan.
2. Terdapat empat aktor yang melakukan kegiatan yaitu staff mainenance, kepala bagian, admin dan system.
3. Terdapat delapan action diantaranya membuat jadwal, melihat jadwal perawatan, membuat laporan perawatan, memeriksa laporan, membuat laporan ganti sparepart, membuat laporan ganti spare part dan mengirim pesan konfirmasi.
4. Satu final node yang merupakan akhir dari kegiatan.

#### C. Class Diagram

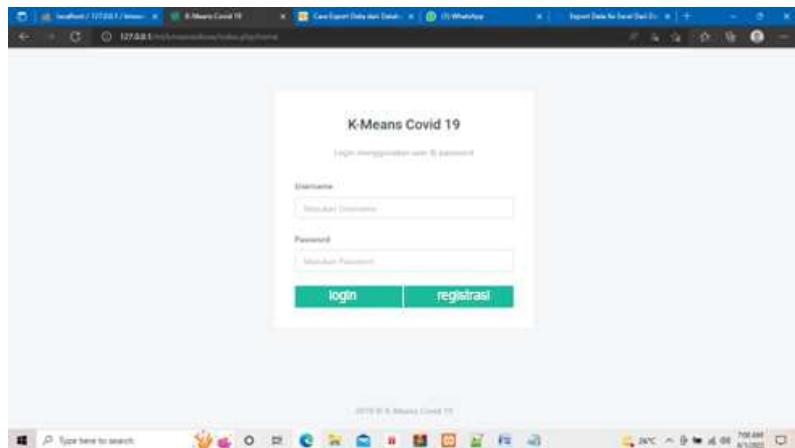
Class Diagram yaitu adalah diagram yang menampilkan setiap class yang saling berhubungan dengan sistem yang dibuat. Berikut class diagram dari aplikasi clustering Covid-19



Gambar 4 Class Diagram

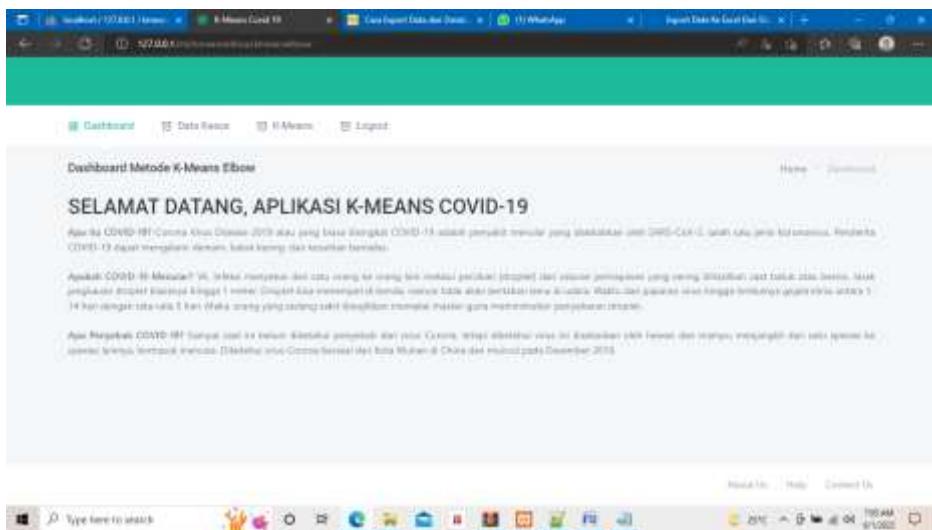
#### D. Desain Aplikasi

Gambar 5 adalah tampilan layar pada halaman login. Saat membuka aplikasi, pengguna langsung diarahkan pada halaman Login. Pada halaman login ini pengguna harus memasukan *username* dan *password* sebelum memulai aplikasi. Jika username dan password salah maka sistem akan membuka halam login kembali. Jika username dan password benar maka sistem akan melanjutkan proses menuju halaman utama.



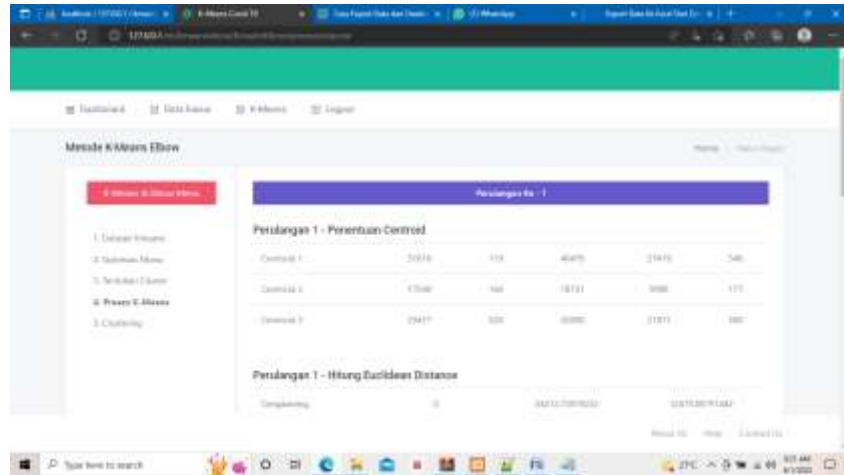
Gambar 5 Tampilan *Login system*

Saat login berhasil, pengguna diarahkan ke halaman Beranda seperti tampilan pada Gambar 6 Tampilan Halaman Utama dibawah ini. Pada halaman ini ditampilkan penjelasan tentang kasus covid-19. Pengguna bisa melanjutkan proses memasukan data dengan memilih menu data kasus untuk lanjut ke proses selanjutnya



Gambar 6 Gambar *desain* tampilan utama

Pada gambar 7 dibawah merupakan halaman yang menampilkan hasil penentuan centroid sebagai berikut. Hasil dari perulangan ke-1 yaitu centroid 1 himpunan 1 berjumlah 3710, himpunan 2 berjumlah 119, himpunan 3 berjumlah 40476, himpunan 4 berjumlah 27419, himpunan 5 berjumlah 548 dan seterusnya sampai centroid 3



Gambar 7 Tampilan Menu Cluster

## V. KESIMPULAN

### A. Kesimpulan

Berdasarkan hasil pengujian dan uji coba pada penerapan metode k-means untuk melakukan pengelompokan atau clusterisasi kasus Covid-19 di DKI Jakarta, dapat disimpulkan bahwa aplikasi clustering metode k-means dapat digunakan dan berjalan dengan baik. jumlah cluster yang direkomendasikan berdasarkan perhitungan nilai SSE adalah 9 cluster. Seperti gambar yang sudah diurutkan dibawah ini berdasarkan perhitungan SSE. Tetapi pada praktiknya, jumlah cluster tersebut bisa ditentukan sendiri oleh pengguna program aplikasi berdasarkan hasil pertimbangan-pertimbangan tertentu. Jika jumlah 9 cluster sangat banyak, penelitian ini merekomendasikan menggunakan 3 (tiga) cluster dikarenakan jumlah tersebut memiliki kualitas yang cukup baik berdasarkan hasil perhitungan metode Elbow dan menempati urutan kedua pada SSE setelah 9 cluster.

Pada clusterisasi dengan 3 (tiga) cluster, diperoleh pengelompokan C1 sebanyak 16 kecamatan, C2 sebanyak 26 kecamatan, dan C3 sebanyak 2 kecamatan. Berdasarkan Indikator Level Asesmen Situasi Tingkat Penularan COVID-19 dan Jumlah kasus COVID-19 yang sudah dicluster dari jumlah penduduk DKI Jakarta per 2021 sebanyak 10.609.681 jiwa, C2 merupakan zona kasus yang tinggi, C1 merupakan zona kasus sedang dan C3 merupakan zona kasus rendah.

### B. Saran

Untuk penelitian selanjutnya, clusterisasi bisa dikembangkan dengan menggunakan berbagai metode clusterisasi yang lainnya seperti Fuzzy K-Means, Fuzzy C-Means, K-Medoids, dan sebagainya dengan mempertimbangkan data yang cocok dan sesuai dan untuk hasil dari clustering K-Means ini bisa dikembangkan lagi agar bisa lebih bermanfaat untuk di bidang lainnya.

## DAFTAR PUSTAKA

- Abdulloh, R. (2015). Web programming is easy. Jakarta: PT. Elex Media Komputindo, 7.
- Bu'ulolo, E., & Purba, B. (2021). Algoritma Clustering Untuk Membentuk Cluster Zona Penyebaran Covid-19. Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi
- Calisher, C. H., Carroll, D., Colwell, R., Corley, R. B., Daszak, P., Drosten, C., & Turner, M. (2021). Science, not speculation, is essential to determine how SARS-CoV-2 reached humans. *The Lancet*, 398(10296), 209-211.
- Florin, G. (2011). Data mining: concepts, models and techniques. Springer-Verlag.
- Haviluddin, H. (2016). Memahami Penggunaan UML (Unified Modelling Language). Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer.

- Jakarta Satu. (2022). Open Data Covid-19 Provinsi DKI Jakarta. <https://riwayat-file-covid-19-dki-jakarta-jakartagis.hub.arcgis.com/>. 1 Januari 2022.
- Keputusan Menteri Kesehatan Republik Indonesia Nomor Hk.01.07/Menkes/413/2020. (2020). Tentang Pedoman Pencegahan Dan Pengendalian Coronavirus Disease 2019 (Covid-19).
- Larose, D. T. (2015). Data mining and predictive analytics
- Madcoms, T. (2016). Pemrograman PHP dan MySQL untuk pemula. Yogyakarta: CV Andi Offset.
- Rahman, A. T. (2017). Coal trade data clustering using K-means (case study Pt. Global Bangkit Utama). ITSMART: Jurnal Teknologi dan Informasi, 6(1), 24-31.
- Santos, T. B. (2019). Aplikasi Data Mining untuk Clustering Daerah Penyebaran Penyakit Diare di DKI Jakarta Menggunakan Algoritma K-MEANS. . J. Ilm. FIFO, 11(2), 131.
- Saputro, R. A., & Widodo, P. P. (2014). Komparasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Penyakit Tuberculosis (Tb): Studi Kasus Puskesmas Karawang Sukabumi.
- Noviyanto, N. (2020). Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia. Paradigma Jurnal Komputer dan Informatika, 183-188.
- Solichin, A., & Khairunnisa, K. (2020). Klasterisasi persebaran virus Corona (Covid-19) di DKI Jakarta menggunakan metode K-Means. Fountain of Informatics Journal, 5(2), 52-59.
- Panjaitan, B., Kom, S., Kom, M., & Turkhamun Adi Kurniawan (2016). CLASTERING DATA AKADEMIK MAHASISWA FAKULTAS TEKNIK USNI DENGAN ALGORITMA K-MEANS.
- Purwasandina, Y., Gunadhi, E., & Fatimah, D. D. S. (2015). Rancang Bangun Sistem Informasi Pengelolaan Data Rujukan Pasien Untuk Puskesmas Pembangunan Garut. Jurnal Algoritma, 12(2), 393-397.
- Suendri, S. (2019). Implementasi Diagram UML (Unified Modelling Language) Pada Perancangan Sistem Informasi Remunerasi Dosen Dengan Database Oracle (Studi Kasus: UIN Sumatera Utara Medan). Algoritma: Jurnal Ilmu Komputer Dan Informatika, 2(2), 1.
- Widodo, P. P., & Prabowo, H. (2011). Menggunakan uml. Bandung:Informatika, 19, 393-403.
- X. Wu et al., "Top 10 algorithms in data mining," in Knowledge and Information Systems, 2008, vol. 14, no. 1, hal. 1–3.