



## INTEGRASI METODE FORWARD SELECTION DAN DISTANCE MEASURE UNTUK MENINGKATKAN AKURASI K-MEANS, X-MEANS DAN K-MEDOID PADA PERSEBARAN DATA COVID-19 DI PROVINSI DKI JAKARTA ( MEI 2021 HINGGA DESEMBER 2021 )

Siti Hotijah<sup>1</sup>, Sudarno Wiharjo<sup>2</sup>, Yan Mitha Djaksana<sup>3</sup>

<sup>1,2,3</sup> Universitas Pamulang

sitihhotijah@gmail.com<sup>1</sup>, sudarnogm@gmail.com<sup>2</sup>, dosen01994@unpam.ac.id<sup>3</sup>

### Kata kunci:

Virus Covid-19, Distance Measure, K-Means, X-Means, K-Medoids, Forward Selection, Davies Bouldin Index

### Abstrak

Virus Covid-19 menarik perhatian global, dimana penambahan jumlah kasus Covid-19 berlangsung cukup cepat serta sudah terjadi penyebaran antar negara. Kota Jakarta merupakan ibu kota Indonesia yang menjadi salah satu pusat pergerakan perekonomian Indonesia sebagai kota terpadat di Indonesia yang menjadi salah satu penyumbang presentase korban terpapar virus covid-19 yang sangat tinggi di Indonesia. Penelitian ini mengintegrasikan metode forward selection dan distance measure untuk meningkatkan akurasi k-means, x-means dan k-medoids pada persebaran data covid-19 di Povinsi DKI Jakarta. Dilakukan percobaan dengan 3 algoritma, 2 distance measure dan penambahan seleksi fitur, kemudian dalam pengolahan data peneliti menggunakan aplikasi Rapidminer Studio. Hasil pengujian dari beberapa metode dan pengukuran jarak dengan tambahan forward selection maupun tanpa forward selection di dapatkan metode terbaik yaitu X-Means menggunakan Euclidean distance. menghasilkan 4 cluster yang kemudian dilakukan uji validasi terhadap hasil cluster yang terbentuk Davies Bouldin Index sebagai metode dalam analisis kluster menghasilkan nilai validitas sebesar 1,132 untuk X-Means tanpa fitur seleksi dan X-Means + Forward Selection dengan Euclidean DBI sebesar 1,336. Yang dapat digunakan sebagai referensi data pola penyebaran virus covid-19 di DKI Jakarta.

### Pendahuluan

Coronavirus adalah salah satu patogen utama yang menjadikan sistem pernapasan manusia, sebagai target utama. Wabah virus corona (CoV) sebelumnya termasuk sindrom pernafasan akut yang parah (SARS)-CoV dan Middle East respiratory syndrome (MERS)-CoV yang sebelumnya pernah terjadi dicirikan sebagai agen yang berarti ancaman kesehatan masyarakat yang besar. Diakhir Desember 2019, sekelompok pasien dirawat di rumah sakit menggunakan diagnosis awal pneumonia dengan etiologi yang tidak diketahui. Sumber penularan kasus ini masih belum diketahui pasti, tetapi kasus pertama dikaitkan dengan pasar ikan di Wuhan.(Rothan & Byrareddy, 2020)

Munculnya 2019-nCoV telah menarik perhatian global, dan pada 30 Januari WHO sudah menyatakan COVID-19 menjadi darurat kesehatan warga yang menjadi perhatian internasional. Penambahan jumlah kasus COVID-19 berlangsung cukup cepat serta sudah terjadi penyebaran antar negara.(Putri, 2020)

Indonesia merupakan negara terpadat keempat di dunia, dan karena itu diprediksi akan sangat menderita dan pada jangka waktu yang lebih lama, jika dibandingkan dengan negara-negara lain yang berpenduduk lebih sedikit. Pada 2 Maret 2020, Presiden Joko Widodo melaporkan dua kasus terkonfirmasi pertama Infeksi COVID-19 di Indonesia.(Djalante et al., 2020)

Kota Jakarta merupakan ibu kota Indonesia yang menjadi salah satu pusat pergerakan perekonomian Indonesia. Salah satu kota terpadat di Indonesia ini memiliki penduduk yang diperkirakan mencapai angka 10,61 juta orang pada tahun 2021, menurut data yang dikeluarkan oleh Survei Penduduk Antar Sensus (SUPAS). Dengan kepadatan penduduk yang sangat tinggi Jakarta menjadi salah satu penyumbang presentase korban terpapar virus covid-19 yang sangat tinggi di Indonesia bersama beberapa kota besar lainnya. Dari penelusuran data terkini Update 11 Desember 2021 diperoleh informasi penduduk di Provinsi DKI Jakarta yang terkonfirmasi positif 864.326 orang, dirawat 124 orang, sembuh 850.475 orang, meninggal 13.583 orang, isolasi mandiri 144 orang, tanpa gejala 42 orang, dan bergejala 31 orang. Dengan data suspek total 5.359.227 orang, probable 33.782 orang, pelaku perjalanan 737.969 orang, kontak erat 1.847.533, discarded 22.556 (Covid-19, n.d.)

Dengan dampak dan banyaknya data yang terus bertambah setiap harinya menunjukkan bahwa terjadi peningkatan jumlah suspect maupun terkonfirmasi virus covid-19 di Jakarta. Untuk itu diperlukan sebuah analisis data yang bisa menghasilkan informasi mengenai pola penyebaran virus covid-19 di Jakarta yang nantinya bisa dijadikan sebagai bahan referensi penanganan virus covid-19. Saat ini data yang ada pada pusat layanan darurat covid-19 DKI Jakarta khususnya pada Link <https://corona.Jakarta.go.id> belum tergambar pola clustering sebagai visualisasi data yang bisa menggambarkan pola penyebaran virus covid-19 di kota Jakarta. Untuk itu diperlukan sebuah analisis data yang bisa menghasilkan informasi mengenai pola penyebaran virus covid-19 di Kota Jakarta yang bisa dijadikan sebagai bahan referensi penanganan virus ini di Kota Jakarta. Serta jika dilihat dari penelitian sebelumnya ada yang pernah menggunakan algoritma k-medoids pada pembentukan cluster zona penyebaran Covid-19 di Sumatra utara oleh (Bu'ulolo & Purba, 2021) penelitian lainnya dengan menganalisis penyebaran penularan Corona di Provinsi Jawa Tengah Menggunakan K-Means Clustering oleh (Darmansah & Wardani, 2021), Berdasarkan penelitian sebelumnya yang membedakan penelitian ini adalah perbandingan tiga algoritma pengelompokan yaitu K-Means, X-Means dan K-Medoids dengan menambahkan 2 pengukuran jarak yaitu Euclidean Distance dan Chebychev Distance, serta dilakukan perbandingan data tanpa fitur seleksi maupun dengan menggunakan fitur seleksi forward selection.

## **Metode**

Peneliti menggunakan dataset german credit data yang bersifat publik (terbuka) dimana data-data yang diperoleh bersumber dari website penanganan covid-19 pemerintah provinsi DKI Jakarta data tersebut terdiri dari 6 wilayah/kota 44 kecamatan dan 267 kelurahan. Selain itu informasi yang terdapat pada media elektronik menjadi salah satu referensi peneliti dalam memahami pergerakan data virus covid-19 di DKI Jakarta yang terus berkembang setiap harinya. Dataset yang diperoleh terdiri dari 13 variabel yaitu

1. Nama Kota.
2. Nama Kelurahan.
3. Suspek.
4. Isolasi dirumah
5. Suspek Meninggal
6. Selesai Isolasi
7. Probable

8. Pelaku Perjalanan
9. Kontak Erat.
10. Positif.
11. Dirawat
12. Sembuh
13. Meninggal

Langkah awal yang peneliti lakukan dalam penelitian ini adalah studi literatur, kegiatan ini dilakukan agar peneliti mampu memahami secara konseptual apa saja yang peneliti harus lakukan dan persiapkan. Dimana data diambil dari jurnal mulai tahun 2019 sampai dengan 2022. Setelah itu langkah yang peneliti lakukan adalah pengumpulan data, dimana data yang digunakan dalam penelitian ini adalah dataset yang di peroleh dari pemerintah provinsi DKI.

Selanjutnya dikarenakan data yang berbentuk tabel belum bisa diolah karena terdapat beberapa kesenjangan nilai data yang terlalu tinggi oleh karena itu data tersebut harus melalui proses normalisasi agar data tersebut bisa menghasilkan hasil yang valid dan sesuai. Selanjutnya dilakukan proses elbow, Metode elbow dapat digunakan untuk menentukan jumlah cluster terbaik yang dapat digunakan untuk menghasilkan hasil cluster terbaik dan dapat memaksimalkan kualitas dari hasil cluster.

Langkah selanjutnya adalah pengolahan data dengan menggunakan dua metode clustering yaitu K-Means, X-Means dan K-Medoids. Untuk mengolah data menjadi cluster-cluster digunakan bantuan aplikasi rapidminer.

Algoritma K-Means merupakan metode yang berusaha mengelompokkan data yang ada menjadi beberapa kelompok, dimana data yang ada dalam satu kelompok mempunyai karakter yang sama satu dengan lainnya dan mempunyai karakter yang berbeda dengan data yang ada dalam kelompok lain.

Algoritma X-Means terdiri dari beberapa operasi berulang hingga eksekusi berakhir, dasar proses dari algoritma x-means adalah memodifikasi struktur cluster k-means dengan membagi data ke dalam minimal dua cluster dilanjutkan dengan pembagian dua kluster untuk setiap kluster yang telah terbentuk

Selanjutnya algoritma K-Medoids yaitu untuk meminimalkan jumlah kesamaan antara setiap objek dan titik referensi yang sesuai. Ditambah menggunakan 2 pengukuran jarak Euclidean Distance dan Chebyshev Distance.

Euclidean Distance adalah panjang ruas garis yang menghubungkan sepasang titik tertentu dalam kisi/grafik, dan merupakan jalur terpendek antar pasangan titik.

Euclidean Distance merupakan salah satu metode perhitungan jarak yang digunakan untuk mengukur jarak dari 2 (dua) buah titik dalam Euclidean space (meliputi bidang euclidean dua dimensi, tiga dimensi, atau bahkan lebih). Untuk mengukur tingkat kemiripan data dengan rumus euclidean distance digunakan rumus berikut (Nishom, 2019):

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan:

dimana,

d = jarak antara x dan y

x = data pusat kluster

y = data pada atribut

i = setiap data

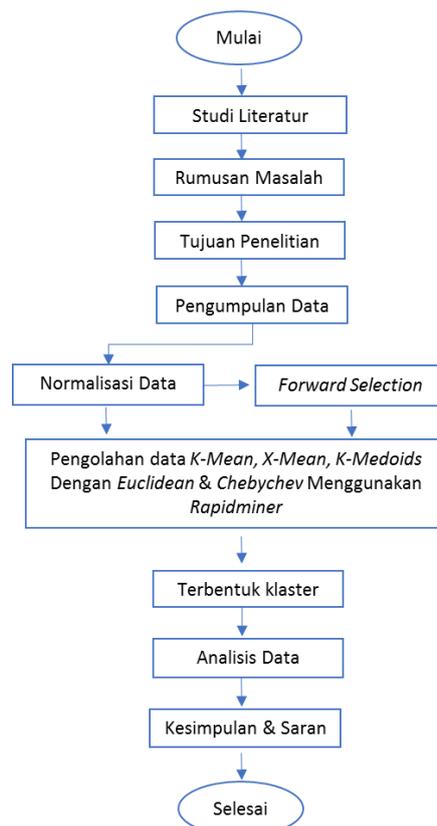
n = jumlah data,

xi = data pada pusat kluster ke i

yi = data pada setiap data ke i

Chebyshev disebut juga maximum distance karena dalam penerapannya mencari jarak terpanjang atau nilai maksimum dari selisih jarak titik data x dan y di dimensi ruang d. Chebyshev didefinisikan sebagai perbedaan terbesar antara dua vector di sepanjang dimensi koordinat apapun, karena Teknik ini hanyalah jarak maksimum dari suatu sumbu.

RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. Dalam hal ini aplikasi ini digunakan untuk mengolah data cluster menggunakan metode K-Means, X-Means dan K-Medoids. Setelah cluster dibentuk dari ketiga metode tersebut maka proses analisis data menjadi langkah selanjutnya yang akan menjadi acuan peneliti dalam menentukan kesimpulan dan saran. Proses validasi cluster dalam penelitian ini menggunakan metode Davies Bouldin Index (DBI). Davies Bouldin Index merupakan salah satu metode pengukuran validitas index dengan memaksimalkan jarak intra-cluster serta disaat yang sama mencoba untuk meminimalisir jarak antar titik dalam sebuah cluster. Jika jarak antar cluster adalah maksimum, perbedaan signifikan pada setiap cluster maka perbedaan kecil antar cluster akan lebih jelas. Jika jarak intra-cluster minimal, berarti setiap objek dalam cluster memiliki tingkat karakteristik yang sangat tinggi. Rancangan penelitian yang akan dilakukan oleh peneliti dapat digambarkan dalam flowChart sebagai berikut:



Gambar 1 Perancangan Penelitian

## Hasil dan Pembahasan

### Penentuan Jumlah Cluster Terbaik

Pada tahapan awal data yang ada akan melalui proses normalisasi data, proses normalisasi data merupakan proses mereduksi perhitungan komputasi yang terlalu besar maka dilakukan proses normalisasi data kedalam range yang lebih kecil yaitu 0.0 s/d 0,2.

Tahap proses selanjutnya dilakukan pada proses Clustering dengan menentukan jumlah K terbaik. Untuk menentukan jumlah K terbaik menggunakan metode Elbow. Pada metode Elbow dalam menentukan jumlah K terbaik dengan melihat fungsi dari nilai cluster pada dataset yang ada yaitu dengan melihat nilai Sum Of Square Error (SSE) pada nilai yang ditentukan dan dibandingkan dengan melihat grafik dari nilai K yang akan di inputkan.

Hasil pengolahan data penelitian untuk penentuan jumlah cluster terbaik menggunakan metode Elbow dengan data sampel berjumlah 267 data. Setelah ditentukan jumlah sampel data dan telah melakukan proses normalisasi data, tahap selanjutnya yang akan dilakukan adalah menentukan berapa jumlah cluster yang tepat untuk digunakan, dimana prosesnya akan dibantu menggunakan aplikasi rapidminer.

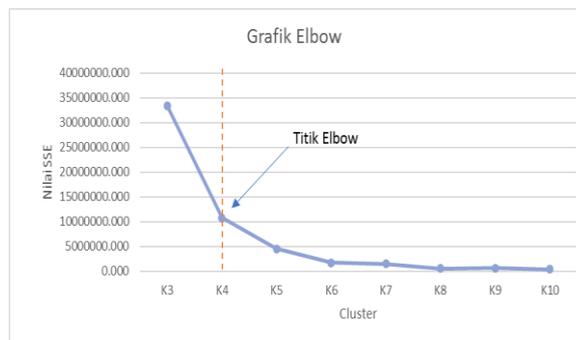
Dari dataset yang ada dengan ujicoba jumlah cluster dari K=2 sampai K=10. Data SSE akan ditampilkan pada table 1. dan kurva elbow akan ditampilkan pada gambar 2, dimana nantinya sudut kurva yang paling siku yang akan digunakan sebagai acuan dalam pemilihan cluster. Selanjutnya digunakan beberapa operator untuk menentukan berapa cluster terbaik yang harus digunakan dan dibutuhkan 6 operator untuk membantu prosesnya seperti Read excel/retrieve, Normalize, multiply, clustering, cluster distance performance.

Setelah data sudah dinormalisasi langkah selanjutnya kita tambahkan multiply agar kita bisa menampilkan /menguji coba beberapa cluster disini uji coba di lakukan dengan menggunakan 9 clustering dan 9 performance. Data proses pencarian cluster akan ditampilkan pada gambar 3.

Table 1 Hasil Sum Square Errorr

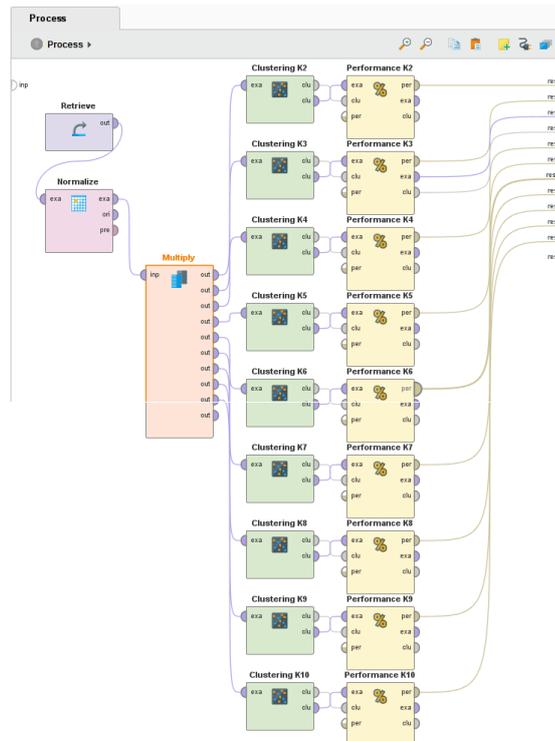
Cluster	Hasil Sum Square Errorr	WCSS
<b>K2</b>	57390388.347	0
<b>K3</b>	23937927.990	33452460.357
<b>K4</b>	13173931.623	10763996.367
<b>K5</b>	8638997.180	4534934.443
<b>K6</b>	6906357.511	1732639.669
<b>K7</b>	5410009.267	1496348.244
<b>K8</b>	4826227.132	583782.135
<b>K9</b>	4196948.497	629278.635
<b>K10</b>	3785166.467	411782.030

Secara visual nilai WCSS atau Within cluster Sum of square dari K yang terdiri dari K2 sampai K10 dapat ditunjukkan dalam gambar 2. di bawah ini:



Gambar 2 Kurva Elbow

Tabel hasil SSE di atas terlihat bahwa selisih antar Sum Square Error pada setiap simulasi cluster diatas K=4 memiliki nilai selisih tertinggi dan hasil visualisasi grafik diatas juga memperlihatkan K4 memiliki sudut siku tertajam di banding K lainnya maka dapat disimpulkan bahwa jumlah cluster terbaik terdapat pada K=3 maka proses pengolahan data clustering dengan metode K-Means, X-Means dan K-medoids akan menggunakan jumlah cluster sebanyak 4 cluster.



Gambar 2 Model Proses Cluster

Tahapan diatas merupakan tahapan dimana dilakukannya proses model cluster dengan parameter jumlah cluster dari 2 klaster hingga 10 klaster. Dengan operator yang digunakan seperti Read excel/retrieve, Normalize, multiply, clustering, cluster distance performance. Berikut penjelasan dari tiap operator yang digunakan:

1. Read Excel

Pada aplikasi rapidminer terdapat fitur read excel yang berfungsi sebagai pembaca file excel. Read excel merupakan operator paling dasar yang digunakan sebelum memulai sebuah proses. Operator ini dapat membaca excel 97,95,2000,2003,xp maupun versi terbaru. File excel yang digunakan berformat.xls.

2. Normalize

Fitur ini digunakan untuk menskala nilai agar sesuai dengan rentang tertentu. Menyesuaikan rentang skala sangat penting Ketika berhadapan dengan atribut dari unit dan skala yang berbeda. Operator ini melakukan normalisasi pada atribut yang dipilih. Dimana dalam penelitian ini rentang yang digunakan untuk normalisasi adalah skala 0,0 sampai 1,0.

3. Multiply

Operator ini digunakan untuk mengambil objek rapiminer dari port input dan mengirimkan salinannya ke port output. Setiap port yang terhubung di dalam multiply membuat salinan independent (tidak terikat).

4. Clustering

Dalam tahap ini dilakukan pengoperasian proses klaster menggunakan algoritma K-Means, X-Means dan K-Medoids.

5. Cluster Distance Performance

Tahapan dalam perform ini bertujuan untuk mencari nilai davies bouldin index. Dimana parameter main criterion disetting davies bouldin dan maximize guna memaksimalkan hasil nilai dari davies bouldin index.

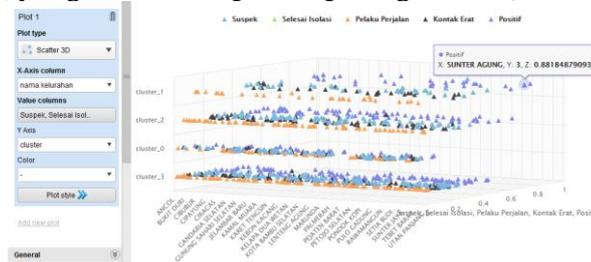
**Analisa Data Tanpa Seleksi Fitur**

Pada tahap awal dilakukan pengujian dengan K-Means Euclidean Dan Chebychev. Hasil cluster data dengan k-means menggunakan jarak euclidean di dapatkan hasil validasi nilai DBI sebesar 1,135 dengan cluster sebanyak 4 cluster dengan bobot nilai sebagai berikut:

- Cluster\_0 dengan jumlah anggota sebanyak 57 dengan persentase 21%
- Cluster\_1 dengan jumlah anggota sebanyak 32 persentase 12%

- Cluster\_2 dengan jumlah anggota sebanyak 72 persentase 27%
- Cluster\_3 dengan jumlah anggota sebanyak 106 persentase 40%.

Dengan informasi tentang titik sebaran kluster per kelurahan menggunakan visualizations dengan tipe Scatter 3D dari hasil proses data yang telah dilakukan. Menunjukkan visualisasi data baru yang terbentuk setelah proses clustering data menggunakan rapidminer, yang akan dilampirkan pada gambar 4.

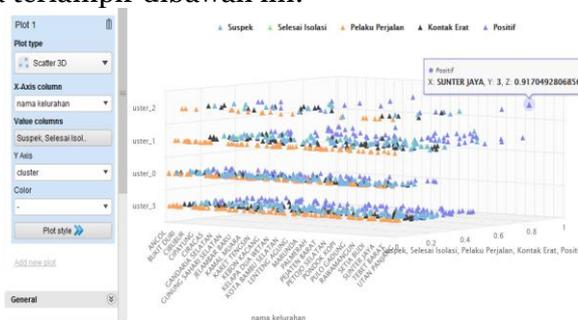


Gambar 3 Visualisasi Hasil *Clustering With K-Means Euclidean*

Selanjutnya hasil dengan k-means dengan jarak Chebychev di dapatkan hasil validasi nilai DBI sebesar 1,378 dengan 4 cluster, dengan bobot nilai sebagai berikut:

- Cluster\_0 dengan jumlah anggota sebanyak 83 dengan persentase 31%
- Cluster\_1 dengan jumlah anggota sebanyak 69 persentase 26%
- Cluster\_2 dengan jumlah anggota sebanyak 18 persentase 7%
- Cluster\_3 dengan jumlah anggota sebanyak 97 persentase 36%.

Dengan visualisasi data terlampir dibawah ini:



Gambar 4 Visualisasi Hasil *Clustering With K-Means Chebychev*

Selanjutnya dilakukan Pengujian Dengan X-Means Euclidean Dan Chebychev. Dengan x-means menggunakan jarak Euclidean di dapatkan hasil validasi nilai DBI 1,132, dengan bobot nilai 4 cluster sebagai berikut:

- Cluster\_0 dengan jumlah anggota sebanyak 32 persentase 12%
- Cluster\_1 dengan jumlah anggota sebanyak 57 dengan persentase 22%
- Cluster\_2 dengan jumlah anggota sebanyak 105 persentase 39%
- Cluster\_3 dengan jumlah anggota sebanyak 73 persentase 27%

Selanjutnya dihasilkan output yang memberikan informasi secara global tentang jumlah data dalam setiap cluster. Dimana diperoleh informasi tentang jumlah kluster data covid per nama kelurahan dan kota yang terbentuk sebagai berikut:

Cluster\_0 : Cengkareng Barat,Cengkareng Timur, Duri Kosambi, Kalideres, Kapuk, Kebon Jeruk, Palmerah, Pegadungan, Tegal Alur, Cilandak Barat, Jagakarsa,Pondok Pinang,Srengseng Sawah, Cibubur, Cipinang Muara, Ciracas, Duren Sawit, Kayu Putih, Klender,Lubang Buaya,Penggilingan, Pondok Bambu, Pondok Kelapa, Pulo Gebang, Pademangan Barat, Pluit, Pademangan Timur, Pegangsaan Dua, Pejagalan,Penjaringan, , Sunter Agung,Sunter Jaya.

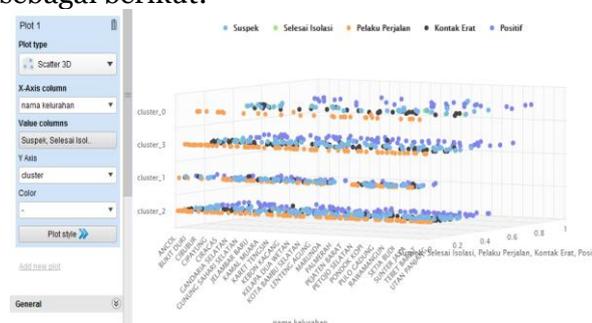
Cluster\_1 : Duri Selatan, Duri Utara, Glodok, Jembatan Besi, Jembatan Lima, Kali Anyar, Keagungan, Kerendang, Krukut, Mangga Besar, Maphar, Pinangsia,Roa Malaka, Galur,Taman Sari, Tambora, Tangki, Bungur, Cikini, kenari, Duri Pulo, Gambir, Gelora, Gondangdia, Kampung Bali, Kampung Rawa, Kebon Kelapa, Kebon Sirih, Kwitang,Petojo

Utara, Senen, Cikoko, Guntur, Karet Semanggi, Karet, Kramat Pela, Kuningan Barat, Melawai, Petogogan, Pulo, Rawa Barat, Setia Budi Selong, Senayan, Bali Mester, Jatinegara Kaum, Kampung Melayu, Rawa Bunga, Rawa Terate, Kamal Muara, Marunda, Pulau Harapan, Pulau Kelapa, Pulau Panggang, Pulau Pari, Pulau Tidung, Pulau Untung Jawa.

Cluster\_2: Angke, Grogol, Jati Pulo, Jelambar Baru, Kamal, Kedaung Kali Angke, Kedoya Selatan, Kelapa Dua, Kembangan Selatan, Kota Bambu Selatan, Slipi, Kota Bambu Utara, Meruya Selatan, Pekojan, Semanan, Sukabumi Utara, Sukabumi Selatan, Tomang, Tanah Sereal, Tanjung Duren Utara, Wijaya Kusuma, Bendungan Hilir, Cideng, Gunung Sahari Selatan, Gunung Sahari Utara, Harapan Mulia, Karang Anyar, Karet Tengsin, Kartini, Kebon Kacang, Kebon Melati, Kemayoran, Kramat, Mangga Dua Selatan, Menteng, Paseban, Pasar Baru, Pegangsaan, Petamburan, Petojo Selatan, Sumur Batu, Tanah Tinggi, Utan Panjang, Bangka, Bukit Duri, Cilindak Timur, Cipete Selatan, Cipete Utara, Cipulir, Duren Tiga, Gandaria Selatan, Gunung, Karet Kuningan, Kebon Baru, Kuningan Timur, Mampang Prapatan, Manggarai, Manggarai Selatan, Menteng Atas, Pancoran, Pasar Manggis, Pengadegan, Pesanggrahan, Petukangan Selatan, Rawa Jati, Tebet Barat, Tebet Timur, Tegal Parang, Bale Kambang, Bambu Apus, Baru, Bidara Cina, Cakung Barat, Cawang, Ceger, Cipinang Besar Selatan, Dukuh, Cipinang Besar Utara, Cipinang Cempedak, Kayu Manis, Kebon Manggis, Munjul, Pal Meriam, Pinang Ranti, Pisangan Baru, Pondok Kopi, Pondok Ronggon, Pulo Gadung, Setu, Ujung Menteng, Utan Kayu Utara, Ancol, Cilincing, Kali Baru, Kapuk Muara, Koja, Papanggo, Rawa Badak Selatan, Rawa Badak Utara, Rorotan, Semper Timur, Sungai Bambu, Tanjung Priok, Tugu selatan, Warakas..

Cluster\_3 : Duri Kupa, Jelambar, Joglo, Kedoya Utara, Kemanggisan, Kembangan Utara, Meruya Utara, Rawa Buaya, Srengseng, Tanjung Duren Selatan, Cempaka Baru, Cempaka Putih Barat, Cempaka Putih Timur, Johar Baru, Kebon Kosong, Rawasari, Serdang, Bintaro, Ciganjur, Cipedak, Gandaria Utara, Grogol Selatan, Grogol Utara, Jati Padang, Kalibata, Kebagusan, Kebayoran Lama Selatan, Kebayoran Lama Utara, Lebak Bulus, Lenteng Agung, Menteng Dalam, Pasar Minggu, Pejaten Barat, Pejaten Timur, Pela Mampang, Petukangan Utara, Pondok Labu, Ragunan, Tanjung Barat, Ulujami, Batu Ampar, Cakung Timur, Cijantung, Cilangkap, Cililitan, Cipayung, Cipinang, Cipinang Melayu, Gedong, Halim Perdana kusumah, Jati, Jatinegara, Kalisari, Kampung Tengah, Kebon Pala, Kelapa Dua Wetan, Kramat Jati, Makasar, Malaka Jaya, Malaka Sari, Pekayon, Pisangan Timur, Rambutan, Rawamangun, Susukan, Utan Kayu Selatan, Kebon Bawang, Kelapa Gading Barat, Kelapa Gading Timur, Lagoa, Semper Barat, Sukapura, Tugu Utara.

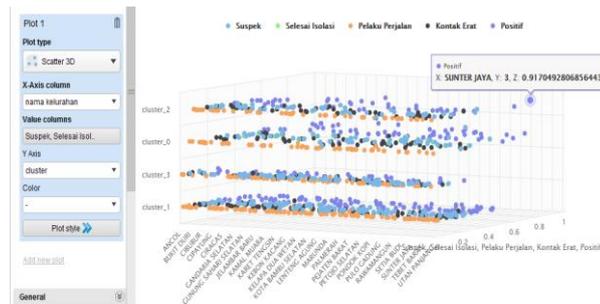
Selanjutnya merupakan hasil proses data yang telah dilakukan menunjukkan visualisasi data baru yang terbentuk sebagai berikut:



Gambar 5 Visualisasi Hasil Clustering *With X-Means Euclidean*

Hasil cluster data dengan x-means menggunakan jarak Chebychev dan hasil visualisasi yang akan di munculkan pada gambar 7, dengan hasil validasi nilai DBI 1,365, dimana di dapatkan bobot nilai sebagai berikut:

- Cluster\_0 dengan jumlah anggota sebanyak 56 persentase 21%
- Cluster\_1 dengan jumlah anggota sebanyak 99 persentase 37%
- Cluster\_2 dengan jumlah anggota sebanyak 46 persentase 17%
- Cluster\_3 dengan jumlah anggota sebanyak 66 dengan persentase 25%

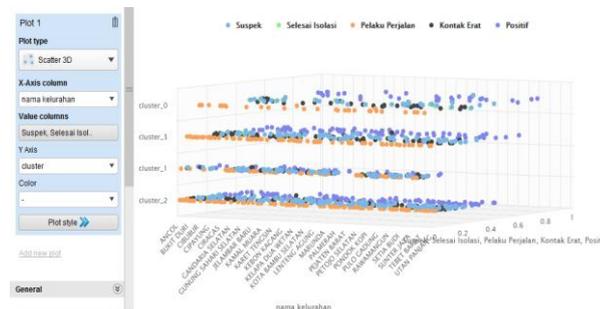


Gambar 6 Visualisasi Hasil *Clustering With X-Means Chebychev*

Selanjutnya dilakukan Pengujian Dengan K-Medoids Euclidean Dan Chebychev. Dengan K-Medoids menggunakan jarak Euclidean di dapatkan hasil validasi nilai DBI 2,187, dengan cluster sebanyak 4 cluster dengan bobot nilai sebagai berikut:

- Cluster\_0 dengan jumlah anggota sebanyak 110 dengan persentase 41%
- Cluster\_1 dengan jumlah anggota sebanyak 38 persentase 14%
- Cluster\_2 dengan jumlah anggota sebanyak 15 dengan persentase 6%
- Cluster\_3 dengan jumlah anggota sebanyak 104 persentase 39%

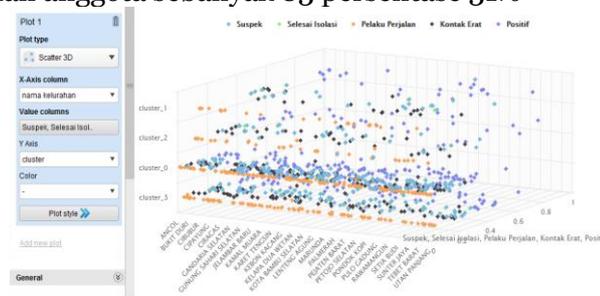
Hasil proses data yang telah dilakukan menunjukkan visualisasi data baru yang terbentuk sebagai berikut:



Gambar 7 Visualisasi Hasil *Clustering With K-Medoids Euclidean*

Hasil cluster data dengan K-Medoids menggunakan jarak Chebychev dengan hasil validasi nilai DBI 1,682 dan hasil visualisasi yang akan di munculkan pada gambar 9, dimana di dapatkan bobot nilai sebagai berikut:

- Cluster\_0 dengan jumlah anggota sebanyak 131 persentase 49%
- Cluster\_1 dengan jumlah anggota sebanyak 22 persentase 8%
- Cluster\_2 dengan jumlah anggota sebanyak 31 persentase 12%
- Cluster\_3 dengan jumlah anggota sebanyak 83 persentase 31%

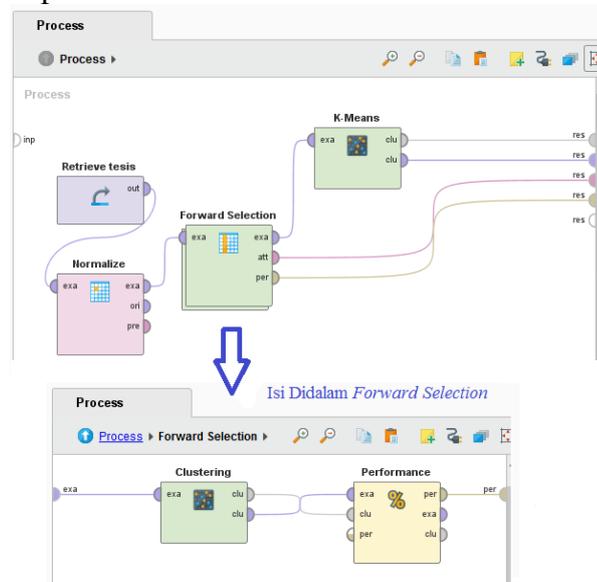


Gambar 8 Visualisasi Hasil *Clustering With K-Medoids Chebychev*

### **Analisa Data Dengan Seleksi Fitur**

Penelitian diawali dengan melakukan pemilihan atribut terbaik dan menghapus atribut yang tidak memiliki kontribusi ke akurasi. Atribut terbaik didapatkan menggunakan metode forward selection. Selanjutnya dilakukan pengukuran jarak yang diawali dengan menggunakan euclidean distance, dilanjutkan dengan menggunakan chebychev distance.

Prosesnya diawali dengan pemanggilan data excel/ retrieve yang dilanjutkan dengan menormalisasi data, lalu menambahkan operator forward selection, didalam operator forward dimasukkan operator clustering dan cluster distance performance, serta dilanjutkan dengan menambah proses cluster diluar. Karena tanpa ada tambahan proses cluster, data yang tampil hanya data atribut serta nilai bobot. Selanjutnya dihasilkan output yang memberikan informasi secara global tentang jumlah data dalam setiap cluster. Dimana prosesnya akan ditampilkan pada Gambar10.



Gambar 10 Proses *Clustering* Menggunakan *Forward Selection*

Pada proses pertama percobaan dilakukan untuk mencari bobot atribut dengan forward selection menggunakan metode k-means. Dihasilkan bobot dengan nilai 1 dan 0, nilai 1 merupakan bobot nilai atribut yang terbaik, yang nantinya akan dijadikan model data. Sedangkan nilai 0 merupakan atribut yang tidak memiliki kontribusi ke akurasi serta akan dihapus demi menghasilkan dataset (baru) yang jumlah atributnya lebih sedikit. Dibawah ini merupakan hasil atau nilai dari pembobotan yang telah diuji dengan menggunakan 2 pengukuran jarak

Table 2 Bobot Atribut *K-Means*

Nama Atribut	Bobot K-Means Euclidean	Nama Atribut	Bobot K-Means Chebychev
Suspek Meninggal	1	Suspek Meninggal	1
Probable	1	Probable	1
Pelaku Perjalan	1	Meninggal	1
Kontak Erat	1	Suspek	1
Dirawat	1	Dirawat	1
Sembuh	1	Pelaku Perjalan	1
Meninggal	1	Sembuh	1
Suspek	0	Selesai Isolasi	1
Isolasi di Rumah	0	Positif	1
Selesai Isolasi	0	Isolasi di Rumah	1
Positif	0	Kontak Erat	0

Table 3 Bobot Atribut X-Means

Nama Atribut	Bobot X-Means Euclidean	Nama Atribut	Bobot X-Means Chebychev
Isolasi di Rumah	1	Suspek Meninggal	1
Suspek Meninggal	1	Probable	1
Probable	1	Positif	1
Kontak Erat	1	Selesai Isolasi	1
Sembuh	1	Dirawat	1
Suspek	0	Meninggal	1
Selesai Isolasi	0	Sembuh	1
Pelaku Perjalan	0	Isolasi di Rumah	1
Positif	0	Pelaku Perjalan	1
Dirawat	0	Suspek	1
Meninggal	0	Kontak Erat	0

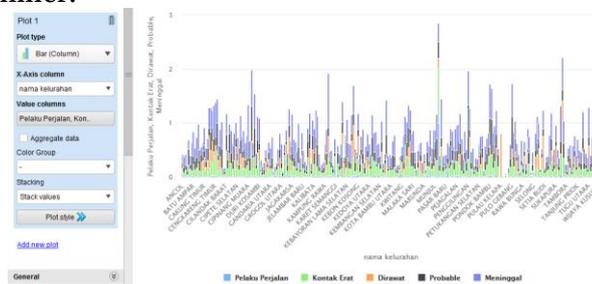
Table 4 Bobot Atribut K-Medoids

Nama Atribut	Bobot K-Medoids Euclidean	Nama Atribut	Bobot K-Medoids Chebychev
Pelaku Perjalan	1	Pelaku Perjalan	1
Kontak Erat	1	Dirawat	1
Dirawat	1	Selesai Isolasi	0
Probable	0	Probable	0
Suspek Meninggal	0	Suspek	0
Selesai Isolasi	0	Meninggal	0
Meninggal	0	Sembuh	0
Sembuh	0	Isolasi di Rumah	0
Suspek	0	Positif	0
Positif	0	Kontak Erat	0
Isolasi di Rumah	0	Suspek Meninggal	0

Dari hasil cluster data setelah proses forward selection data dengan k-means menggunakan jarak Euclidean didapatkan hasil validasi nilai DBI 1,342, dimana didapatkan cluster sebanyak 4 cluster dengan banyaknya anggota sebagai berikut:

- Cluster\_0, anggota sebanyak 114 persentase 43%
- Cluster\_1, anggota sebanyak 52 persentase 19%
- Cluster\_2, anggota sebanyak 74 persentase 28%
- Cluster\_3, anggota sebanyak 27 persentase 10%

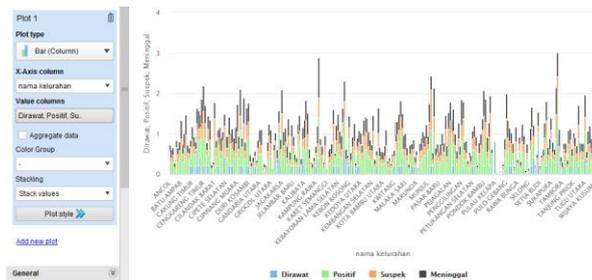
Selanjutnya dapat dilihat dibawah ini merupakan hasil proses data yang telah dilakukan yang menunjukkan visualisasi data baru yang terbentuk setelah proses clustering data menggunakan rapidminer.



Gambar 11 Visualisasi Hasil Clustering With K-Means Euclidean

Selanjutnya dilakukan pengujian dengan k-means menggunakan jarak chebychev di dapatkan didapatkan hasil validasi nilai DBI 1,524, dimana dihasilkan cluster sebanyak 4 cluster dengan bobot nilai sebagai berikut:

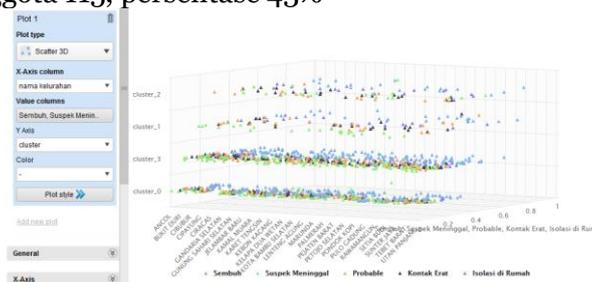
- Cluster\_0 jumlah anggota 57, persentase 21%
- Cluster\_1 jumlah anggota 79, persentase 30%
- Cluster\_2 jumlah anggota 41, persentase 15%
- Cluster\_3 jumlah anggota 90, persentase 34%



Gambar 129 Visualisasi Hasil *Clustering With K-Means Chebychev*

Dari hasil cluster data setelah proses forward selection data dengan x-means menggunakan jarak Euclidean didapatkan hasil validasi nilai DBI 1,336, dimana didapatkan cluster sebanyak 4 cluster dengan banyaknya anggota sebagai berikut:

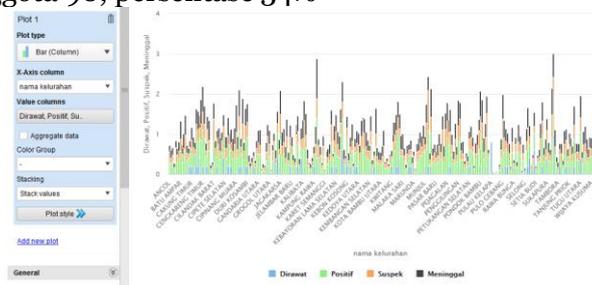
- Cluster\_0 jumlah anggota 110, persentase 41%
- Cluster\_1 jumlah anggota 23, persentase 9%
- Cluster\_2 jumlah anggota 19, persentase 7%
- Cluster\_3 jumlah anggota 115, persentase 43%



Gambar 13 Visualisasi Hasil *Clustering With X-Means Euclidean*

Selanjutnya dilakukan pengujian dengan k-means menggunakan jarak chebychev di dapatkan didapatkan hasil validasi nilai DBI 1,524, dimana dihasilkan cluster sebanyak 4 cluster dengan bobot nilai sebagai berikut:

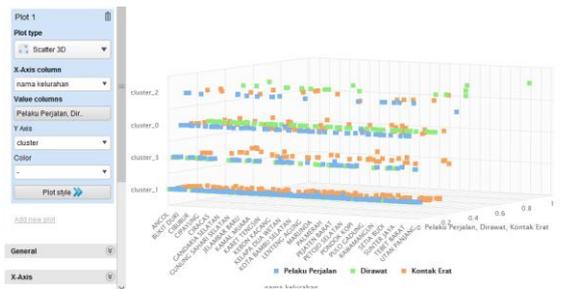
- Cluster\_0 jumlah anggota 57, persentase 21%
- Cluster\_1 jumlah anggota 79, persentase 30%
- Cluster\_2 jumlah anggota 41, persentase 15%
- Cluster\_3 jumlah anggota 90, persentase 34%



Gambar 14 Visualisasi Hasil *Clustering With X-Means Chebychev*

Dari hasil cluster data setelah proses forward selection data dengan k-medoids menggunakan jarak Euclidean didapatkan hasil validasi nilai DBI 6,669, dimana didapatkan cluster sebanyak 4 cluster dengan banyaknya anggota sebagai berikut:

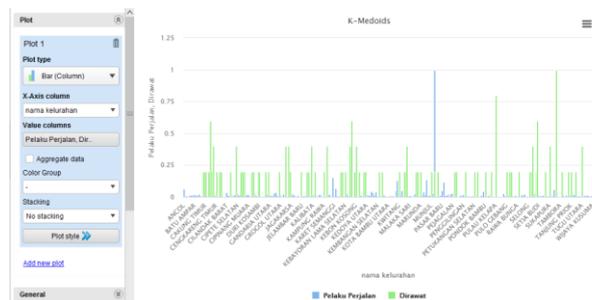
- Cluster\_0 jumlah anggota 62, persentase 23%
- Cluster\_1 jumlah anggota 145, persentase 54%
- Cluster\_2 jumlah anggota 17, persentase 7%
- Cluster\_3 jumlah anggota 43, persentase 16%



Gambar 15 Visualisasi Hasil *Clustering With K-Medoids Euclidean*

Selanjutnya dilakukan pengujian dengan k-means menggunakan jarak chebychev di dapatkan didapatkan hasil validasi nilai DBI 7,346, dimana dihasilkan cluster sebanyak 4 cluster dengan bobot nilai sebagai berikut:

- Cluster\_0 jumlah anggota 117, persentase 56%
- Cluster\_1 jumlah anggota 17, persentase 8%
- Cluster\_2 jumlah anggota 61, persentase 29%
- Cluster\_3 jumlah anggota 15, persentase 7%



Gambar 16 Visualisasi Hasil *Clustering With K-Medoids Chebychev*

## Pembahasan

Tujuan dari penelitian ini adalah membandingkan tiga algoritma pengelompokan dengan melakukan penyeleksian atribut maupun tanpa seleksi atribut. Algoritma yang digunakan yaitu K-Means, X-Means dan K-Medoids ditambah dengan penggunaan pengukuran jarak euclidean distance dan chebychev distance. Dimana data akan divalidasi menggunakan Davies Bouldin Index.

Berdasarkan hasil pengujian data akan dibagi menjadi 4 klaster. Data awal yang diuji yaitu pengujian tanpa fitur seleksi dilanjutkan menggunakan tambahan fitur seleksi yaitu forward selection.

Selanjutnya didapatkan hasil dari perhitungan nilai DBI yang akan digunakan sebagai perbandingan efektifitas clustering ketiga metode yang digunakan yaitu K-Means, X-Means dan K-Medoids. Semakin kecil nilai Davies Bouldin Index (DBI) yang diperoleh (Non Negatif  $\geq 0$ ), maka akan semakin baik cluster yang dihasilkan dari pengelompokan menggunakan ketiga metode tersebut.

Table 5 Hasil *Davies Bouldin Index (DBI)*

<b>Pengujian Tanpa Seleksi</b>		
<b>Algoritma</b>	<b>Nilai Davies Bouldin Index</b>	
	<b>Euclidean Distance</b>	<b>Chebychev Distance</b>
<b>K-Means</b>	1,135	1,378
<b>X-Means</b>	1,132	1,365
<b>K-Medoids</b>	2,187	1,682
<b>Pengujian Dengan Forward Selection</b>		
<b>Algoritma</b>	<b>Nilai Davies Bouldin Index</b>	
	<b>Euclidean Distance</b>	<b>Chebychev Distance</b>

<b>K-Means+Forward Selection</b>	1,342	1,524
<b>X-Means+Forward Selection</b>	1,336	1,524
<b>K-Medoids+ Forward Selection</b>	6,669	7,346

## Kesimpulan

Setelah melakukan perhitungan dan pengolahan data dengan menggunakan metode K-Means, X-Means dan K-Medoids dengan maupun tanpa forward selection dalam pengklasteran data covid-19 di wilayah DKI Jakarta dapat disimpulkan sebagai berikut:

Pola penyebaran virus covid-19 di wilayah DKI Jakarta dikelompokkan menjadi 4 kluster terbaik. Dengan hasil validasi nilai Davies Bouldin Index terbaik tanpa fitur seleksi adalah dengan menggunakan metode X-Means dengan nilai DBI 1,132 dan distance measure terbaik menggunakan euclidean. Sedangkan hasil validasi nilai Davies Bouldin Index terbaik dengan fitur seleksi forward selection adalah dengan menggunakan metode X-Means dengan nilai 1,336 dan distance measure terbaik menggunakan euclidean.

## Daftar Pustaka

- Bu'ulolo, E., & Purba, B. (2021). Algoritma Clustering Untuk Membentuk Cluster Zona Penyebaran Covid-19. *Jurnal Teknologi Informasi & Komunikasi*, 12(1), 59–67.
- Covid-19. (n.d.). Retrieved December 11, 2021, from <https://corona.jakarta.go.id/id/data-pemantauan>
- Darmansah, & Wardani, N. W. (2021). Analisis Pesebaran Penularan Virus Corona di Provinsi Jawa Tengah Menggunakan Metode K-Means Clustering. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(1), 105–117. <https://doi.org/10.35957/jatisi.v8i1.590>
- Djalante, R., Lassa, J., Setiamarga, D., Sudjatma, A., Indrawan, M., Haryanto, B., Mahfud, C., Sinapoy, M. S., Djalante, S., Rafliana, I., Gunawan, L. A., Surtiari, G. A. K., & Warsilah, H. (2020). Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020. *Progress in Disaster Science*, 6. <https://doi.org/10.1016/j.pdisas.2020.100091>
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 20–24. <https://doi.org/10.30591/jpit.v4i1.1253>
- Putri, R. N. (2020). Indonesia dalam Menghadapi Pandemi Covid-19. *Jurnal Ilmiah Universitas Batanghari Jambi*, 20(2), 705. <https://doi.org/10.33087/jiubj.v20i2.1010>
- Rothan, H. A., & Byraredy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of Autoimmunity*, 109(February), 102433. <https://doi.org/10.1016/j.jaut.2020.102433>