JURNAL INFORMATIKA, Vol.9 No.1 April 2022, Halaman 43-54
ISSN: 2355-6579 | E-ISSN**:** 2528-2247

Analisis Penyebaran *Pandemi Covid-19* di Kota Jakarta Menggunakan Metode *Clustering K-Means* dan *Density Based Spatial Clustering of Application With Noise*

Tukiyat[1], Yohanes Djohan[2]

[1,2] Universitas Pamulang
Jl. Surya Kencana No.1 Pamulang Barat, Kecamatan Pamulang, Kota Tangerang Selatan, Indonesia

e-mail: [1] tukiyat@bppt.go.id, [2]anezjo123@gmail.com

Informasi Artikel	Diterima: 20-08-2021	Direvisi: 24-12-2021	Disetujui: 29-12-2021

Abstrak

Pandemi Covid-19 awal Maret 2020 telah masuk di Indoensai dan penyebaran virus *covid-19* di Jakarta sudah sangat mengkhawatirkan, dengan wilayah yang cukup luas Jakarta menjadi salah satu daerah dengan perkembangan warga terinfeksi virus *covid-19* yang sangat tinggi. Penelitian ini bertujuan untuk menganalisis pola penyebaran virus *covid-19* di kota DKI Jakarta sehingga mampu memberikan gambaran *cluster* umum yang menjadi pusat pergerakan virus ini di wilayah DKI Jakarta. Data penelitian berupa data sekunder yang diperoleh dari sumber https://corona.Jakarta.go.id. Berjumlah 267 data. Sampel penelitian sebanyak sebanyak 151 data. Data tersebut selanjutnya akan dianalisis untuk memetakan pola penyebaran virus *covid-19* di kota Jakarta dengan menggunakan metode *K-Means* dan *Density Based Spatial Clustering of Applications With Noise* (DBSCAN). Hasil analisis metode *K-Means* menunjukan bahwa pandemi virus *covid-19* terbagi dalam empat *cluster* yaitu *cluster* 0 dengan anggota sebanyak 91 data, *cluster* 1 dengan anggota sebanyak 23 data, *cluster* 2 sebanyak 62 data dan *cluster* 3 sebanyak 91 data. Sedangkan untuk metode DBSCAN menghasilkan empat *cluster* dengan komposisi *cluster* 0 sebanyak 87 data, *cluster* 1 sebanyak 23 data, *cluster* 2 sebanyak 63 data dan *cluster* 3 sebanyak 94 data. Hasil validasi *cluster* menggunakan *Davies Bouldin Index* menunjukan bahwa dalam penelitian ini metode *K-Means* lebih baik dibanding metode DBSCAN dengan perbandingan *index* DBI 0,027968805 untuk *K-Means* dan 1,172562165 untuk DBSCAN.

Kata Kunci: *Covid-19, K-Means, Clustering, Density Based Clustering*.

Abstract

The Covid-19 pandemic in early March 2020 has entered Indoensai and the spread of the covid-19 virus in Jakarta is already very worrying, with a large area of Jakarta being one of the areas with the development of people infected with the covid-19 virus is very high. This study aims to analyze the pattern of the spread of the covid-19 virus in the city of DKI Jakarta so that it is able to provide an overview of the general cluster that is the center of this virus movement in the DKI Jakarta area. The research data is in the form of secondary data obtained from the source https://corona.Jakarta.go.id. Total 267 data. The research sample consisted of 151 data. The data will then be analyzed to map the pattern of the spread of the Covid-19 virus in the city of Jakarta using the K-Means method and Density Based Spatial Clustering of Applications With Noise (DBSCAN). The results of the K-Means method analysis show that the Covid-19 virus pandemic is divided into four clusters, namely cluster 0 with 91 data members, cluster 1 with 23 data members, cluster 2 with 62 data and cluster 3 with 91 data. Meanwhile, the DBSCAN method produces four clusters with the composition of cluster 0 as many as 87 data, cluster 1 as many as 23 data, cluster 2 as many as 63 data and cluster 3 as many as 94 data. The results of cluster validation using the Davies Bouldin Index show that in this study the K-Means method is better than the DBSCAN method with a DBI index ratio of 0.027968805 for K-Means and 1.172562165 for DBSCAN.

Keywords*: Covid-19, K-Means, Clustering, Density Based Clustering*

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

1. Pendahuluan

Saat ini Virus *Covid-19* berkembang dengan sangat pesat hampir diseluruh dunia, menurut Organisasi Kesehatan Dunia (WHO) virus ini bertransmisi melalui *droplet* saat seseorang batuk, bersin, bernyayi, berbicara hingga bernafas dimana pada proses tersebut udara yang keluar dari hidung dan mulut mengeluarkan partikel kecil atau *aerosol* dalam jarak dekat. Cara lain virus tersebut menyebar yaitu melalui media permukaan benda yang terkontaminasi virus *covid-19* dari orang yang positif terpapar dari proses batuk ataupun bersin dimana virus ini bisa bertahan selama 2-3 hari dipermukaan benda tertentu.

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 10,57 juta orang pada tahun 2020, 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. Tercatat pertanggal 09 September 2020 kasus terkonfirmasi *Covid-19* di kota Jakarta sudah mencapai jumlah 49.837 ribu jiwa jika dibandingkan dengan jumlah kasus terkonfirmasi *Covid-19* secara nasional yaitu berjumlah 200.342 ribu jiwa kota Jakarta sudah memberikan sumbangsih sebesar lebih dari 24 %, data ini diambil dari tanggal 4 maret 2020 sampai tanggal 09 September 2020 melalui *Link* Layanan Darurat *Covid-19* Kota DKI Jakarta.

Tingginya *trend* peningkatan kasus terko nfir-masi virus *Covid-19* di kota Jakarta saat ini sangat mengkhawatirkan, untuk itu maka diperlukan kerja keras dan kerja sama oleh semua elemen yang ada untuk meminimalisir ataupun mencegah berke mba-ng nya virus ini secara cepat dan tepat. Dengan data yang berkembang setiap harinya menunjukan bahwa terjadi peningkatan jumlah *suspect* maupun terkonfirmasi virus ini di daerah ibu kota Jakarta dalam beberapa bulan terakhir, untuk itu diperlukan sumbangsih dari berbagai pihak yang ada untuk memberikan manfaat pencegahan berkembangnya virus ini, salah satunya adalah dengan menganalisa data yang ada menjadi sebuah informasi yang mampu memberikan gambaran pola penyebaran virus *covid-19* di Kota DKI Jakarta.

2. Metode Penelitian

Populasi merupakan keseluruhan obyek yang berada pada suatu wilayah dan memenuhi syarat-syarat tertentu berkaitan dengan masalah penelitian, atau keseluruhan unit atau individu yang akan diteliti. Adapun yang menjadi popula si di dalam penelitian ini adalah populasi tak terbatas warga DKI Jakarta yang terkait dengan virus *covid-19* di wilayah DKI Jakarta selama kurun waktu dimulainya wabah ini hingga tanggal 09 September 2020, data tersebut didapatkan dari website penanganan *covid-19* https://corona.Jakarta.go.id dimana dalam data yang didapat, populasi ini terdiri dari warga DKI Jakarta yang terdiri dari 45 kecamatan dan 269 kelurahan yang terkait dengan *virus covid-19* berjumlah 267 orang yang diambil dari tanggal 4 maret 2020 sampai tanggal 09 September 2020. Sampel penelitian berjumlah 149 data yang diambil dengan metode *Stephen Issac* dan *William B Michael* dengan Rumus :

$$S= \frac{\lambda^2 .N .P.Q}{d^2 (N-1)+ \lambda^2 .PQ}$$

Dimana :
S = Jumlah *Sample.*
N = Jumlah Populasi.
$\boldsymbol{\lambda^2}$ = *Chi kuadrat* (Tergantung derajat kebebasan dan tingkat kesalahan. Untuk derajat kebebasan 1 dan kesalahan 5 % harga *chi kuadrat* = 3,841).
d = Perbedaan antara *sampel* yang diharapkan dengan yang terjadi perbedaan. Perbedaan bisa 1%, 5% dan 10 %.
P = Q = 0,5.

Jumlah *dataset* yang peneliti dapatkan berjumlah 267 data dengan menggunakan *koefisien* d = 5%, maka jumlah sampel yang digunakan dalam penelitian ini adalah sebanyak 151 data yang didapatkan dari perhitungan dengan metode *interpolasi*.

Dari hasil studi *literatur* yang peneliti lakukan terhadap beberapa penelitian dan karya tulis ilmiah, terdapat beberapa literasi pengetahuan yang memiliki keterkaitan dengan penelitian yang peneliti lakukan.

Pada penelitian yang ditulis oleh (Bastian et al., 2018) dari Majalengka *University* dengan judul "Penerapan Algoritma *K-Means Clustering Analysis* Pada Penyakit Menular Manusia", dimana penggunaan metode *K-Means* di titik beratkan pada pengolahan data set *variabel* yang dibentuk per kecamatan yang berjumlah 32 puskesmas di kabupaten Majalengka. Analisis yang dilakukan berhasil mengelompokan penyakit menular manusia berdasarkan frekuensi kejadian. Dengan meng*cluster*kan penyakit menular pada manusia di wilayah Majalengka penulis berhasil membentuk *cluster Homogenitas internal* dan *Heterogenitas eksternal* yang cukup tinggi, sehingga pola

penyebaran penyakit menular di wilayah tersebut bisa dikenali.

Penelitian lain yang bisa menggambarkan penggunaan metode *K-Means* yaitu pada penelitian yang dilakukan oleh (Pratiwi et al., 2015) dengan judul "Penerapan *K-Means Clustering* Untuk Memprediksi Minat Nasabah Pada PT. Asuransi Jiwa Bersama 1912 Bumi Putera Prabumulih" dimana dalam penilitian ini penulis menggunakan tahapan *Knowledge Discovery in Database* (KDD) yang terdiri dari *Data Cleaning, data Integration, Data Selection, Data Transformation, Data Mining, Evaluation dan Presetation* yang berhasil mengelompokan minat yang cukup tinggi untuk menjadi nasabah asuransi dengan asuransi terpilih adalah mitra Beasiswa pada calon nasabah yang berprofesi sebagai petani.

Pada judul "Pemilihan *Distance Measure* Pada *K-Means Clustering* Untuk Pengelompokan Member Di Alvaro Fitnes" karya (Anggara et al., 2016) penggunaan *K-Means clustering* digunakan untuk mengelompokan member di Alvaro *Fitnes* digunakan pengujian dengan metode *Silhouette Coefficient* yang menghasilkan *distance measure* yang paling optimal dengan nilai yang paling mendekati 1 adalah 0,242821. Dalam hal ini *distance measure* digunakan untuk mengukur kemiripan data dalam suatu *cluster*.

Sedangkan pada karya tulis berjudul "Penerapan Algoritma *K-Means* Untuk *Clustering* Data Obat-Obatan Pada RSUD Pekanbaru" oleh (Gustientiedina et al., 2019) menggunakan Algoritma *K-Means* sebagai metode untuk men*gcluster*kan perencanaan kebutuhan obat-obatan yang akan menghasilkan pengadaan obat-obatan menjadi efektif dan efisien sehingga data *cluster* yang dihasilkan bisa digunakan dalam refernsi pengambilan keputusan dalam perencanaan dan pengendalian pasokan medis di rumah sakit tersebut.

Untuk metode DBSCAN ada beberapa karya tulis yang mampu menambah khasanah pengetahuan penulis dalam menggunakan metode DBSCAN untuk teknik *clustering*. Penelitian tersebut antara lain terdapat pada Karya Tulis (Aritonang et al., 2015) dari Universitas Sumatera Utara yang berjudul "Analisis *Subspace Clustering* Menggunakan DBSCAN dan SUBLU Untuk Proyeksi Pekerjaan Alumni Perguruan Tinggi", dimana pada tulisan ini penulis menggunakan metode DBSCAN dan SUBCLU untuk mengelompokan data atau atribut pada *cluster* yang berbeda. Dengan menentukan tingkat kerapatan data dan juga mengindentifikasi *Outlier* atau data yang tidak relevan. Hasil akhir yang didapatkan bahwa SUBCLU tidak memiliki *un-cluster* dataset yang nyata, sehingga peresepsi hasil *cluster* akan menghasilkan informasi yang lebih akurat dan DBSCAN memerlukan waktu lebih lama dari pada metode SUBCLU sehingga kinerja SUBCLU lebih efisien daripada DBSCAN.

Sementara itu dalam karya tulis "Perbandingan Metode *K-Means* Dengan Metode DBSCAN Pada Pengelompokan Rumah Kost Mahasiswa Di Kelurahan Tembalang Semarang" yang ditulis oleh (Budiman et al., 2016) dari Institut Universitas Diponegoro, dimana pada karya tulis ini dibandingkan dua buah metode peng*clusteran* yaitu metode *K-Means* dan DBSCAN dalam mengelompokan rumah kost mahasiswa didaerah Semarang yang menghasilkan kesimpulan bahwa metode *K-Means* bekerja lebih baik di banding DBSCAN dalam mengelompokan rumah kost mahasiswa yang dibuktikan dengan nilai indeks *silhoutte* pada *K-Means* sebesar 0.463 lebih tinggi dari nilai indeks *Silhoutte* pada DBSCAN yaitu sebesar 0.281.

Dalam studi literatur penulis mengenai penggunaan metode DBSAN ini salah satu Jurnal yang menarik untuk dipelajari adalah karya tulis (Sari & Primajaya, 2019) dengan judul *"* Penerapan *Clustering* DBSCAN Untuk Pertanian Padi Di Kabupaten Karawang", dimana pada penerapan metode DBSCAN dilakukan pada lahan pertanian padi di kabupaten karawang dengan memperhatikan nilai dari *average silhoutte width* untuk menunjukan kualitas *cluster* yang terbentuk. Kesimpulan yang dihasilkan dari karya tulis ini adalah penerapan metode *clustering* DBSCAN pada penelitian ini adalah dua *cluster* yang memiliki perbedaan karakteristik yang dipengaruhi oleh tingkat curah hujan, luas lahan, serangan hama dll.

Sebagai objek dari penelitian ini virus *covid-19* juga termasuk dalam daftar referensi penulis untuk mengetahui penelitian-penelitian apa saja yang sudah dilakukan terkait dengan virus *covid-19* ini, dimana hal ini dimaksudkan agar penulis memahami pola pergerakan dan segala hal terkait dengan virus ini sebagai bahan referensi penulis dalam menyelesaikan tulisan ini. Pada karya tulis yang berjudul "Pandemi *Covid-19*, Respon Imun Tubuh. Dan *Herd Imunity*" karya (Handayani et al., 2020) dimana tujuan dari penulisan ini yaitu untuk memberikan informasi mengenai definisi pandemi*,* sistem imunitas tubuh dalam menghadapi *covid-19* dijelaskan bahwa pandemi *covid-19* terjadi karena ada penemuan dan mutasi baru dari virus *Sars-CoV* menjadi sangat infeksius dan *virulensi* tinggi. Dimana pada kesimpulan akhir penulis menyimpulkan bahwa waktu yang sesuai sangat penting dalam pendeteksian penyebaran *covid-19*. Untuk menghindari bias dan hasil

negatif palsu, seluruh tenaga kesehatan dan mayoritas masyarakat harus memahami sistem imunitas tubuh dan mekanisme alamiah virus di dalam tubuh.

Dalam penggunaan metode *Clustering* terutama pada metode *K-Means clustering* penggunaan jumlah *cluster* ditentukan pada saat awal pembentukan *cluster*, untuk itu salah satu hal yang peneliti harus pelajari secara medalam yaitu bagaimana menentukan jumlah *cluster* terbaik pada metode *K-Means Clustering*. Dari beberapa referensi yang peneliti temukan karya tulis dari (Putu et al., 2016) dari Universitas Atmajaya dengan judul " Analisa Penentuan Jumlah *Cluster* Terbaik Pada Metode *K-Means Clustering*" sangat membantu peneliti dalam memahami bagaimana menemukan jumlah *cluster* terbaik dalam tulisan ini. Dalam tulisannya penulis menggunakan metode *Elbow* dalam menentukan jumlah *cluster* terbaik. Metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik. Dengan memilih nilai *cluster* dan menambah nilai *cluster* tersebut untuk dijadikan model data dalam penentuan jumlah *cluster t*erbaik. Hasil yang didapatkan menyimpulkan bahwa Penentuan jumlah *cluster t*erbaik dengan metode *elbow* dapat menghasilkan jumlah c*luster K* yang sama pada jumlah data yang berbeda-beda..

Perbandingan jumlah *cluster* terbaik pada *K-Means* juga dapat di lihat pada karya tulis berjudul "Perbandingan *Distance Space Manhattan* Dengan *Euclidian* Pada *K-Means Cluste*ring Dalam menentukan Promosi" karya (Fajriah et al., 2019) dimana dalam penelitian ini penulis membandingkan dua *distance space* yaitu *manhattan* dan *euclidian* untuk menentukan jumlah *cluster* terbaik yang menghasilkan pernyataan bahwa metode *euclidian distance* merupakan metode paling optimal dalam penentuan jumlah *cluster* terbaik dibandingkan dengan mentode *manhattan* dengan perbandingan lebih kecil yaitu 15,115 < 15,398.

Referensi lain yang merujuk pada penggunaan metode untuk menentukan jumlah *cluster* terbaik terdapat pada karya tulis (Rahman et al., 2017) dengan judul "*Coal Trade* Data *Cluste*ring Using *K-Means* (*Case Study* PT. Global Bangkit Utama)" yang menggunakan metode *Elbow* untuk mendapatkan jumlah *cluster* yang optimal berdasarkan evaluasi nilai SSE maka didapatkan jumlah *cluster* sebanyak 7 merupakan *cluster* terbaik pada penelitian ini.

Data yang digunakan dalam penelitian ini di dapatkan dari *website* penanganan *covid-19* pemerintah daerah DKI Jakarta pada *Link https://corona.*Jakarta*.go.id.* dimana data tersebut berbentuk *dataset* yang terdiri dari 45 wilayah kecamatan dengan 269 wilayah kelurahan. Selain itu informasi yang terdapat pada media *elektronik* maupun cetak menjadi salah satu referensi peneliti dalam memahami pergerakan data virus *covid-19* di DKI Jakarta yang terus berkembang setiap harinya.

Penelitian dirancang dengan pendekatan *dekriptif* yang mampu mendeskripsikan atau memberi gambaran terhadap objek yang diteliti melalui data atau sampel yang telah didapatkan. Dengan menggunakan penelitian *deskriptif* maka data-data yang telah dikumpulkan, diolah, dianalisis dan disajikan secara sistematis akan mampu mendeskripsikan objek-objek yang peneliti dapatkan dan bersifat apa adanya.

Pada penulisan tesis ini penulis mendesain penelitian dengan metode *kuantitatif* dengan perencanaan dalam penelitian yang terperinci dengan menjabarkan hubungan antar *variabel* penelitian yang terdapat pada *dataset* yang didapatkan dari *Link* https://corona.Jakarta.go.id untuk menguji teori serta melakukan *generalisasi* fenomena sosial mengenai penyebaran *covid-19* di DKI Jakarta.

Metode pendekatan yang digunakan dalam penelitian ini adalah *Unsupervised Learning* dimana *Unsupervised Learning* digambarkan sebagai pendekatan yang tidak menggunakan label sebagai dasar untuk membuat prediksi *Variable,* melainkan menggunakan kesamaan attribut dari attribut yang dimiliki. Dengan menggunakan *Unsupervised Learning* maka dataset yang digunakan oleh peneliti bisa secara langsung mengelompokan data kedalam *cluster* tanpa perlu menggunakan data latih atau data *training* sehingga proses *clustering* bisa dilakukan dengan lebih cepat.

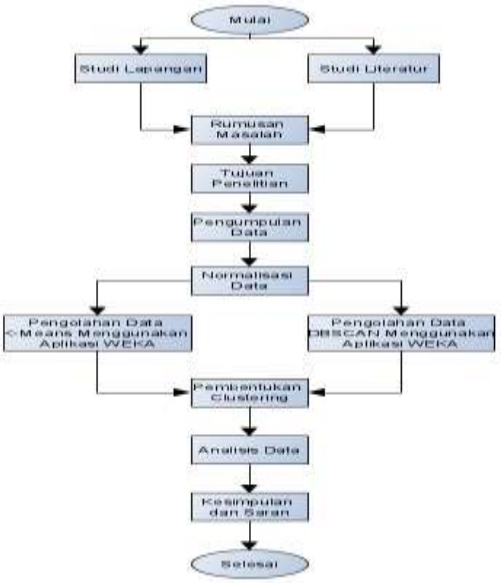
Jenis data yang digunakan dalam penelitian ini adalah data *kuantitatif.* Dimana data yang diperoleh peneliti dalam penelitian ini bisa diukur dan dihitung secara langsung dalam bentuk bilangan atau angka. Data-data yang diperoleh dari *Link* https://corona.Jakarta.go.id merupakan kumpulan data yang disajikan dalam bentuk bilangan atau angka, dalam hal ini penulis bermaksud menyajikan data dalam olahan data *mining* untuk menghasilkan presentasi data yang membentuk *cluster* penyebaran *covid-19* di Kota DKI Jakarta.

Langkah awal yang peneliti lakukan dalam penelitian ini adalah studi *literatur* dan studi lapangan, kegiatan ini dilakukan agar peneliti mampu memahami secara *konseptual* apa saja

yang peneliti harus lakukan dan persiapkan dalam menyelesaikan tesis ini. Setelah mendapatkan referensi *literatur* maka langkah selanjutnya yang peneliti lakukan adalah dengan membuat rumusan masalah dan tujuan penelitian berdasarkan studi *literatur* dan studi lapangan yang telah peneliti lakukan sebelumnya. Setelah itu langkah yang peneliti lakukan adalah pengumpulan data dimana data yang digunakan dalam penelitian ini adalah *dataset* dari *link* https://corona.Jakarta.go.id data yang berbentuk tabel tersebut 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.

Langkah selanjutnya adalah pengolahan data dengan menggunakan dua metode *clustering* yaitu *K-Means* dan DBSCAN. Untuk mengolah data menjadi *cluster-cluster* digunakan bantuan *aplikasi WEKA*. *Aplikasi WEKA* sendiri merupakan *aplikasi* yang mampu mengolah data secara tepat dan cepat menjadi data-data yang sudah terstruktur. Dalam hal ini *aplikasi* ini digunakan untuk mengolah data *cluster* menggunakan metode *K-Means* dan DBSCAN. Setelah *cluster* dibentuk dari kedua 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.

Perencanaan pada tahap kegiatan yang akan dilakukan oleh peneliti dapat dilihat dari gambar 1.



Gambar 1. *FlowChart* Proses Perancangan Penelitian.

Hasil analisis data selanjutnya dianalisis dengan statistik *deskriptif* dari *dataset* untuk menggambarkan karakteristik penyebaran data dan analisis *cluster*ing dengan metode *K-Means* dan DBSCAN.

Prosedur analisis *cluster*ing menggunakan metode *K-Means* langkah yang digunakan adalah sebagai berikut :

1. Tentukan jumlah *K* sebagai jumlah *cluster* yang akan di input (Jumlah *K* didapat dari hasil perhitungan menggunakan metode *Elbow*).
2. Tentukan *k Centroid* awal secara random.
   $v = \sum^{n}_{i=1} / n\ x_i$; i= 1,2,3,...*n*
   Dimana ; v : *Centroid* pada *cluster*
   x_i : Objek ke – *i*
   n : Banyaknya objek/jumlah objek yang menjadi anggota *cluster.*
3. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing *cluster*.
4. Untuk menghitung jarak antara objek dengan *centroid* dapat menggunakan *Euclidian Distance* :
   $d(x,y) = |x-y|| = \sqrt{\sum^{n}_{i=1}(x_i-y_i)^2}$; $i = 1,2,3,...,n$
   Dimana ; x_i : Objek x ke-i
   y_i : Daya y ke-i
   n : Banyaknya Objek
5. Alokasikan masing-masing objek ke dalam *centroid* yang terdekat.
6. Lakukan iterasi kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan (1).
7. Ulangi langkah 3 jika posisi *centroid* baru tidak sama.

Sedangkan pada proses DBSCAN langkah-langkah yang dilakukan adalah sebagai berikut :

1. Inisialisasi
   1.1. Bobot W*ft* (acak).

1.2. Laju Pemahaman awal dan faktor penurunannya.
1.3. Bentuk dan jari-jari (=R) topologi sekitarnya.
2. Selama kondisi penghentian nilai salah, lakukan langkah 3-7.
3. Untuk setiap *vektor* masukan x, lakukan langkah 4-6.
4. Hitung $D_j = \sum_t (W_{ft}\text{-}X_t)^2$ untuk semua j
5. Tentukan indeks J sedemikian hingga $D_{(J)}$ minimum.
6. Untuk setiap unit j di sekitar J modifikasikan bobot :
   $W_{ft}^{baru} = W_{ft}^{lama} + \alpha\ (X_t - W_{ft}^{lama})$
7. Modifikasi laju pemahaman.
8. Uji kondisi penghentian.

Pada analisis *dataset* yang diperoleh digunakan fitur data *analysis* pada *Ms. Excel* 2010, dimana fitur ini berfungsi untuk memetakan penyebaran data dan juga statistik data yang ada. Terlihat pada tabel 3.3. bahwa analisis data yang dilakukan menghasilkan beberapa point data seperti pada tabel di atas.

Untuk menguji validasi metode *cluster*ing digunakan *metode Davies Bouldin Index* (DBI), yaitu dengan menghitung rata-rata nilai setiap titik pada himpunan data dan jumlah cluster terbaik ditunjukan dengan nilai DBI yang semakin kecil. Dalam menghitung nilai DBI diperlukan langkah-langkah sebagai berikut :

1. Menentukan nilai *Sum of Square Within Cluster* (SSW) dengan persamaan sebagai berikut.

$$SSW_i = \frac{1}{m_i} \sum_{i=i}^{m_i} d\,(x_j, c_i)$$

2. Menentukan nilai *Sum of Square Between Cluster* dengan persamaan sebagai berikut.

$$SSB_{i,j} = d\,(c_i, c_j)$$

3. Menentukan nilai *Ratio* dengan persamaan sebagai berikut.

$$R\ i,j = \frac{1}{mi}\ ssw_i$$

4. Maka persamaan dari ketiga rumus tersebut diatas akan digunakan untuk mencari nilai index DBI dengan menggunakan persamaan berikut.

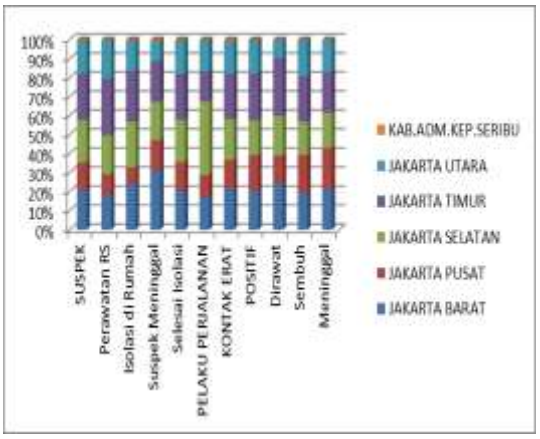
$$DBI = \frac{1}{k} \sum_{i=1}^{k} max\,(R_{i,j})$$

Kriteria uji model *clustering* yang baik adalah model yang memiliki nilai DBI terendah. semakin rendah nilai index DBI maka *cluster*ing akan semakin baik, begitu pula sebaliknya apabila nilai DBI besar maka tingkat validasi atau akurasi dalam *clustering* semakin kurang baik.

3. Hasil Dan Pembahasan

Penyebaran kasus *covid-19* di Kota Jakarta dari analisis data dapat dideskripsikan bahwa kasus Suspek sebanyak 85.238 jiwa, kasus selesai isolasi sebesar 76.098 jiwa, dirawat sebanyak 1.155 jiwa, perawatan sebanyak 1.132 jiwa, pelaku perjalanan sebanyak 2.012 jiwa, kasus sembuh sebanyak 35.425 jiwa, isolasi di rumah sebanyak 7.536 jiwa, kontak erat sebanyak 154.419 jiwa, meninggal sebanyak 1.158 jiwa, suspek meninggal sebanyak 517 jiwa dan kasus positif sebanyak 45.601 jiwa.

Dari deskripsi tersebut maka variabel jumlah kasus covid 19 yang paling besar adalah kontak erat dengan jumlah 154.419 jiwa dengan persentase sebesar 38 % dari kasus yang terjadi dalam variable – variable yang terkait dengan penyebaran virus covid-19 di daerah DKI Jakarta.



Gambar 2. Grafik Kasus *Covid-19* di DKI Jakarta.

Gambar 2 menunjukan komposisi setiap daerah di Jakarta yang menyumbang kasus *covid-19* secara menyeluruh, terlihat pada grafik tersebut setiap daerah di DKI Jakarta memiliki *range* area yang hampir sama, terkecuali daerah kepulauan seribu yang hampir tidak terlihat dalan grafik tersebut. Hal ini menunjukan bahwa penyebaran kasus covid yang merata di setiap daerah DKI Jakarta yang berada dalam satu daratan yang sama, sedangkan untuk wilayah Kepulauan Seribu kasus *covid-19* masih dalam kondisi yang aman.

Pada pengolahan *Dataset* dengan metode *K-Means* dan DBSCAN ini peneliti menggunakan bantuan dari aplikasi WEKA untuk menentukan jumlah *cluster* optimal dalam proses *cluster*ing ini dengan menggunakan metode *elbow*. Dalam pengolahan data kali ini penulis menggunakan fitur *Clustering* yang terdapat pada aplikasi weka untuk menghasilkan *cluster-cluster* dari dataset yang ada. 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 dengan menggunakan aplikasi WEKA untuk me*normalisasi*kan data.

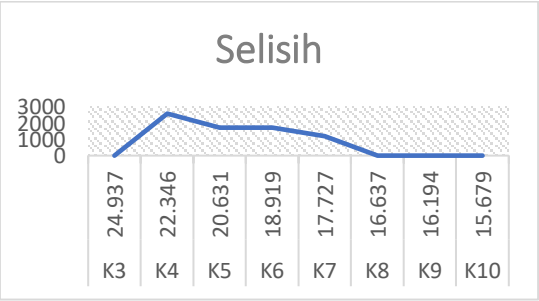
Tahap selanjutnya yang akan dilakukan dalam melakukan proses *K-Means Clustering* adalah menentukan jumlah *K* terbaik yang akan digunakan dalam penelitian ini. Untuk menentukan jumlah *K* terbaik penulis menggunakan metode *elbow* dalam menentukan jumlah *K* terbaik dimana metode ini akan melihat fungsi dari nilai *cluster* pada *dataset* yang ada, dengan melihat nilai *Sum Of Square Erorr* (SSE) pada nilai yang ditentukan dan dibandingkan dengan melihat grafik dari nilai *K* yang akan di inputkan.

Proses penentuan jumlah *cluster* terbaik menggunakan metode *Elbow* dengan data sampel berjumlah 151 data, sampel yang digunakan adalah urutan 20 sampai 171 dari *dataset* yang ada dengan ujicoba jumlah *cluster* dari K=3 sampai K=10.

Tabel 1. Hasil SSE.

Cluster	Jumlah Data	Hasil Sum Square Erorr	Selisih
K3	151	24.937	-
K4	151	22.346	2.591
K5	151	20.631	1.715
K6	151	18.919	1.712
K7	151	17.727	1.192
K8	151	16.637	1.09
K9	151	16.194	0.443
K10	151	15679	0.515

Tabel 1 menunjukan nilai selisih dari K yang terdiri dari K0 sampai K10.

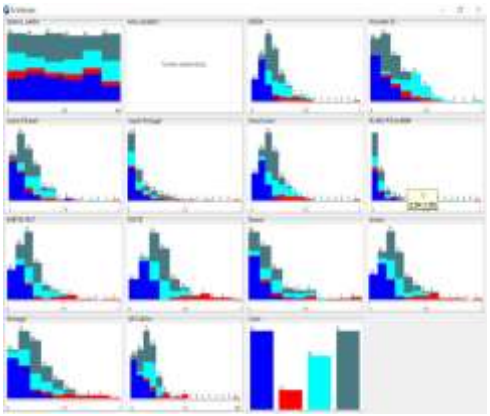


Gambar 3. Grafik Nilai SSE.

Gambar 3 memperlihatkan 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=4* maka proses pengolahan data *clustering* dengan metode *K-Means* dan DBSCAN akan menggunakan jumlah *cluster* sebanyak 4 *cluster*.

Langkah awal dalam membentuk *cluster* pada aplikasi WEKA adalah dengan menginputkan dataset kedalam fitur *cluster* dimana jumlah *cluster* yang akan di bentuk adalah berjumlah *K*=4 sesuai dengan hasil penentuan jumlah *cluster* terbaik menggunakan metode *Elbow*.

Proses *clustering* pada aplikasi *WEKA* diatas menghasilkan 4 *cluster* dengan komposisi *cluster* 0 sebesar 34 %, *cluster* 1 sebesar 9 %, *cluster* 2 sebesar 23 % dan *cluster* 3 sebesar 34 %.

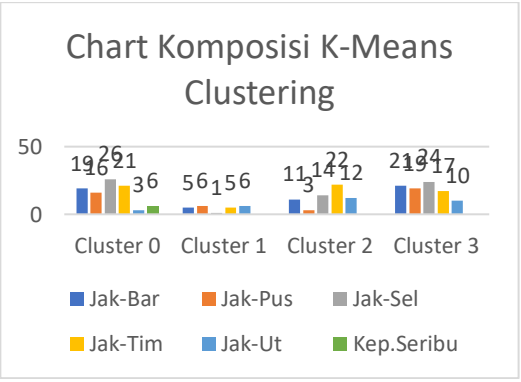


Gambar 4. Visualisasi Data *Cluster*ing Per Atribut.

Gambar 4 memperlihatkan komposisi *cluster* terhadap semua atribut yang ada yang masing masing atribut memberikan kontribusi yang berbeda-beda pada *cluster* yang terbentuk.

Dari hasil pengolahan data diatas dihasilkan *cluster* sebanyak 4 dengan bobot nilai sebagai berikut:

Cluster 0 dengan jumlah anggota sebanyak 91 dengan *presentase* 34 %.
Cluster 1 dengan jumlah anggota sebanyak 23 dengan *presentase* 9 %.
Cluster 2 dengan jumlah anggota sebanyak 62 dengan *presentase* 23 %.
Cluster 3 dengan jumlah anggota sebanyak 91 dengan *presentase* 34 %.



Gambar 5. *Chart* Komposisi *Clustering K-Means*.

Tabel 2. Kontribusi *Cluster K-Means* setiap daerah.

	Jak-Bar	Jak-Pus	Jak-Sel	Jak-Tim	Jak-Ut	Kep. Seribu
C0	19	16	26	21	3	6
C1	5	6	1	5	6	0
C2	11	3	14	22	12	0
C3	21	19	24	17	10	0
Jumlah	56	44	65	65	31	6

Pada tabel 2 terlihat kontribusi tertinggi pada daerah Jakarta Barat ada pada *cluster* 3 dengan jumlah 21 begitu juga dengan daerah Jakarta Pusat yang memiliki kontribusi tertinggi pada *cluster* 3 dengan nilai 19, sedangkan pada daerah Jakarta Selatan kontribusi terbesar ada pada *cluster* 0 dengan nilai 26. Untuk daerah Jakarta Timur dan Jakarta Utara *cluster* 2 merupakan kontribusi tertinggi dengan Jakarta Timur bernilai 22 dan Jakarta Utara dengan nilai 12. Untuk daerah Kepulauan Seribu hanya berkontribusi di *cluster* 0 dengan nilai 6.

Tabel 3. Kontribusi Persentase *Cluster K-Means* Setiap Daerah.

	Jak-Bar	Jak-Pus	Jak-Sel	Jak-Tim	Jak-Ut	Kep. Seribu
C0	19.21 %	16.18 %	26.28 %	21.23 %	3.30%	6.70%
C1	22%	26%	4%	22%	26%	0%
C2	18%	5%	23%	35%	19%	0%
C3	23%	21%	26%	19%	11%	0%
Jumlah	63%	52%	53%	76%	59.3 %	6.70%

Tabel 3 menunjukan kontribusi *presentase* setiap daerah pada semua *cluster* yang terbentuk maka daerah Jakarta Timur merupakan daerah dengan persentase tertinggi dengan persentase sebesar 76 % di ikuti dengan daerah Jakarta Barat dengan persentase sebesar 63 % lalu Jakarta Utara dengan persentase sebesar 59 % dan di peringkat berikutnya adalah Jakarta Selatan dengan persentase sebesar 53 % serta Jakarta Pusat di posisi berikutnya dengan persentase sebesar 52 %, sedangkan khusus untuk daerah Kepulauan Seribu di posisi terakhir dengan persentase keseluruhan berjumlah 6,70 %.

Dilihat dari letak geografis dan luas wilayah dimana daerah Jakarta Barat seluas 126,15 km2, Jakarta Pusat dengan luas 47,90 km2, Jakarta Selatan dengan luas 145,73 km2, Jakarta Timur memiliki luas 187,73 km2, Jakarta Utara dengan luas 142,20 km2 dan Kepulauan Seribu dengan luas daerah 11,81 km2 (Sumber https://jakarta.go.id) maka daerah Jakarta Timur merupakan daerah yang sangat tertinggi mengalami kasus *covid-19* di wilayah Jakarta hal ini terkait dengan luas wilayah Jakarta Timur yang memang secara luas daerah merupakan daerah terluas di wilayah DKI Jakarta.

Tabel 4. Peringkat *K-Means* Kontribusi Setiap Daerah.

	Persentase	Luas Daerah/km2	Peringkat
Jak-Tim	**76%**	**187,73**	**1**
Jak-Bar	**63%**	**126,15**	**2**
Jak-Ut	**59.3 %**	**142,2**	**3**
Jak-Sel	**53%**	**145,73**	**4**
Jak-Pus	**52%**	**47,9**	**5**
Kep. Seribu	**6,70%**	**11,81**	**6**

Dilihat dari Tabel 4. diatas pada peringkat 1 daerah Jakarta Timur memang memiliki daerah paling luas di antara daerah lainnya di Kota DKI Jakarta, akan tetapi pada peringkat 2 yang di tempati oleh daerah Jakarta Barat memiliki daerah yang luasnya 126,15 km2 tidak seluas daerah Jakarta Utara dan Jakarta Selatan pada peringkat 3 dan 4 yang masing-masing memiliki luas daerah sebesar 142.2 dan 145,73 km2. Hal ini menunjukan bahwa penyebaran *covid-19* di daerah Jakarta Barat lebih tinggi di banding daerah lainnya di kota DKI Jakarta. Begitu juga dengan daerah Jakarta Utara yang memiliki luas daerah sebesar 142,2 km2 masih memiliki persentase yang lebih tinggi dari daerah Jakarta Selatan yang memiliki luas wilayah sebesar 145,73 km2.

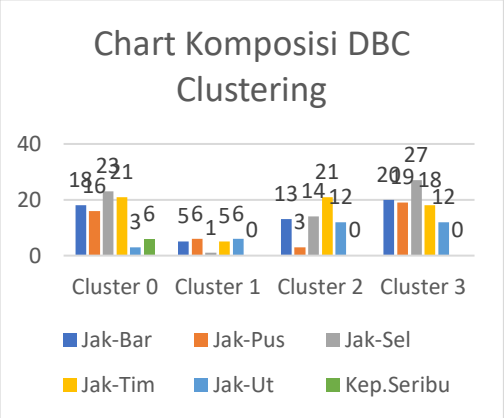
Khusus untuk wilayah Kepulauan Seribu yang memiliki luas daerah terkecil memang tidak memberikan dampak yang sangat signifikan dalam penelitian ini, hal ini bisa disebabkan oleh letak geografis daerah kepulauan seribu yang memang terisolasi oleh lautan dari wilayah DKI Jakarta lainnya, sehingga dampak penyebaran *covid-19* di daerah ini masih sangat kecil. Dari paparan diatas maka dapat disimpulkan bahwa luas area di daerah DKI Jakarta bukan merupakan salah satu faktor yang paling berpengaruh dalam penyebaran virus *covid-19* di wilayah DKI Jakarta.

Seperti pada proses *cluster*ing *K-Means* jumlah *cluster* yang digunakan dalam metode ini adalah *K*=4 sesuai dengan hasil penentuan jumlah *cluster* terbaik menggunkaan metode *Elbow*.

Proses *clustering* menggunakan metode DBSCAN pada aplikasi *WEKA* diatas menghasilkan 4 *cluster* dengan komposisi *cluster* 0 sebesar 33 %, *cluster* 1 sebesar 9 %, *cluster* 2 sebesar 24 % dan *cluster* 3 sebesar 35 %.

Dari hasil pengolahan data mengunakan metode DBSCAN diatas dihasilkan *cluster* sebanyak 4 dengan bobot nilai sebagai berikut :
Cluster 0 dengan jumlah anggota sebanyak 87 dengan *presentase* 33 %.
Cluster 1 dengan jumlah anggota sebanyak 23 dengan *presentase* 9 %.

Cluster 2 dengan jumlah anggota sebanyak 63 dengan *presentase* 24 %.
Cluster 3 dengan jumlah anggota sebanyak 94 dengan *presentase* 35 %.



Gambar 6. *Chart* Komposisi *Clustering* DBSCAN.

Pada Gambar 6 memperlihatkan sebuah komposisi sumbangan setiap daerah pada *cluster* yang terbentuk dimana setiap warna yang berbeda mewakili setiap daerah kota di Jakarta.

Tabel 5. Kontribusi *Cluster* DBSCAN Setiap Daerah.

	Jak-Bar	Jak-Pus	Jak-Sel	Jak-Tim	Jak-Ut	Kep. Seribu
C0	18	16	23	21	3	6
C1	5	6	1	5	6	0
C2	13	3	14	21	12	0
C3	20	19	27	18	10	0
Jumlah	56	44	65	65	31	6

Tabel 5 memperlihatkan kontribusi tertinggi daerah Jakarta Barat pada *cluster* yang terbentuk ada pada *cluster* 3 dengan jumlah 20 begitu juga dengan daerah Jakarta Pusat yang memiliki kontribusi tertinggi pada *cluster* 3 dengan nilai 19, pada daerah Jakarta Selatan kontribusi terbesar ada pada *cluster* 3 dengan nilai 27. Untuk daerah Jakarta Timur pada *cluster* 0 dan *cluster* 2 dengan nilai 21, Jakarta Utara *cluster* 2 dengan nilai 12. Untuk daerah Kepulauan Seribu hanya berkontribusi di *cluster* 0 dengan nilai 6.

Tabel 6. Kontribusi Persentase *Cluster* DBSCAN Setiap Daerah.

	Jak-Bar	Jak-Pus	Jak-Sel	Jak-Tim	Jak-Ut	Kep. Seribu
C0	21%	18%	26%	24%	4%	7%
C1	22%	26%	4%	22%	26%	0%
C2	21%	5%	22%	33%	19%	0%
C3	22%	20%	29%	18%	11%	0%
Jumlah	86%	69%	81%	60%	60%	7%

Jika dilihat dari dari tabel 6 diatas kontribusi jumlah *presentase* setiap daerah pada semua *cluster* yang terbentuk maka daerah Jakarta Timur merupakan daerah dengan persentase tertinggi dengan persentase sebesar 97 % di ikuti dengan daerah Jakarta Barat dengan persentase sebesar 86 % lalu Jakarta Selatan dengan persentase sebesar 81 % dan di peringkat berikutnya adalah Jakarta Pusat dengan persentase sebesar 69 % serta Jakarta Utara di posisi berikutnya dengan persentase sebesar 60 %, sedangkan khusus untuk daerah Kepulauan Seribu di posisi terakhir dengan persentase keseluruhan berjumlah 7 %.

Tabel 7. Peringkat DBSCAN Kontribusi Setiap Daerah.

	Persentase	Luas Daerah/km2	Peringkat
Jak-Tim	97%	187,73	1
Jak-Bar	86%	126,15	2
Jak-Sel	81%	145,73	3
Jak-Pus	69%	47,9	4
Jak-Ut	60%	142,2	5
Kep. Seribu	7,00%	11,81	6

Pada tabel 7 kontribusi persentase *cluster* setiap daerah terlihat daerah Jakarta Timur merupakan daerah dengan persentase tertinggi dengan 97 % dari total sumbangan anggota *cluster* di setiap *cluster* yang terbentuk diikuti dengan daerah Jakarta Barat sebesar 86 % lalu Jakarta selatan di posisi berikutnya dengan nilai persentase sebesar 81 % dan wilayah Jakarta pusat di peringkat selanjutnya dengan nilai persentase sebesar 69 % diikuti dengan wilayah Jakarta Utara dengan nilai persentase 60 % dan diperingkat terakhir wilayah Kepulauan Seribu dengan nilai persentase 7 %.

Jika dilihat dari luas wilayah pada masing-masing daerah memang wilayah Jakarta Timur memiliki area yang paling luas yaitu sebesar 187.76 km2, sedangkan di peringkat kedua wilayah Jakarta Barat masih lebih kecil dari peringkat ketiga yaitu Jakarta Selatan dengan luas wilayah 145,73 km2. Hal ini menunjukan bahwa luas wilayah belum bisa menjadi acuan dalam menentukan seberapa besar kasus penyebaran virus *covid-19* di wilayah DKI Jakarta. Begitu juga dengan wilayah Jakarta Utara yang memiliki luas wilayah 142,2 km2 masih di bawah wilayah Jakarta Pusat yang hanya memiliki luas wilayah 47,9 km2.

Tabel 8. Perbandingan Hasil *Cluster*ing *K-Means* dan DBSCAN

K-Means				
Cluster	C0	C1	C2	C3
Jumlah Data	91	23	62	91

DBSCAN				
Cluster	C0	C1	C2	C3
Jumlah Data	87	23	63	94

Pada tabel 8 memperlihatkan perbandingan dari hasil *cluster*ing pada metode *K-Means* dan DBSCAN terlihat pada tabel diatas. Dengan jumlah data yang terdapat beberapa perbedaan jumlah data yang ada di *cluster* 0, *cluster* 2 dan *cluster* 3. Selisih pebedaan yang tidak terlalu tinggi ini ternyata menghasilkan perubahan peringkat yang signifikan hal ini terlihat pada tabel 9. dibawah ini.

Tabel 9. Perbandingan Peringkat Kontribusi *Cluster*ing.

K-Means				DBSCAN			
	Persentase	Luas Daerah/Km2	Peringkat		Persentase	Luas Daerah/Km2	Peringkat
Jak-Tim	97%	187.73	1	Jak-Tim	76%	187.73	1
Jak-Bar	86%	126.15	2	Jak-Bar	63%	126.15	2
Jak-Sel	81%	145.73	3	Jak-Ut	59%	142.2	3
Jak-Pus	69%	47.9	4	Jak-Sel	53%	145.73	4
Jak-Ut	60%	142.2	5	Jak-Pus	52%	47.9	5
Kep. Seribu	7%	11.81	6	Kep. Seribu	6.70%	11.81	6

Terdapat perubahan dari peringkat ke tiga sampai peringkat ke lima. Pada pengolahan data menggunakan metode *K-Means* wilayah Jakarta Selatan berada di peringkat ke tiga akan tetapi pada pengolahan data menggunakan metode DBSCAN wilayah Jakarta Selatan turun peringkat menjadi posisi ke empat, begitu juga dengan posisi ke empat dan kelima dimana wilayah Jakarta pusat dari peringkat ke empat turun menjadi peringkat ke lima dan wilayah Jakarta Utara yang sebelumnya berada di peringkat ke lima malah naik keperingkat ke tiga.

Jika dilihat dari luas daerah yang di miliki oleh setiap wilayah di kota DKI Jakarta ternyata tidak mempengaruhi kontribusi penyebaran virus *covid-19* di wilayah DKI Jakarta, hal ini bisa jadi di akibatkan oleh perbedaan tingkat kepadatan penduduk, tingkat migrasi para pekerja maupun faktor lainnya. Gambaran umum tingkat kepadatan penduduk di wilayah Administrasi DKI Jakarta dapat dilihat pada tabel dibawah ini yang di ambil dari website https://data.Jakarta.go.id/.

Tabel 10. Kepadatan Penduduk Di Diwilayah DKI Jakarta.

Tahun	Kab/Kota_Administrasi	Luas_Area	Jumlah_penduduk	Kepadatan_Penduduk
2014	Jakarta Timur	188.03	2,817,994	14967.83
	Jakarta Barat	129.54	2,430,410	18761.85
	Jakarta Selatan	141.27	2,164,070	15318.68
	Jakarta Utara	146.66	1,729,444	11792
	Jakarta Pusat	48.13	910.381	18915.04
	Kepulauan Seribu	8.7	23.011	2644943

Jika dilihat dari tabel 10 diatas wilayah Jakarta Timur baik dari aspek luas wilayah dan jumlah penduduk memang memiliki jumlah yang lebih tinggi di banding wilayah lainnya, begitu juga dengan wilayah Jakarta Barat yang memiliki jumlah penduduk terbanyak kedua setelah wilayah Jakarta Timur. Hal ini sesuai dengan hasil *cluster*ing yang dilakukan dengan metode *K-Means* dan DBSCAN dimana kedua wilayah tersebut tetap konstan di peringkat pertama dan kedua. Sedangkan untuk tiga wilayah lainnya selain wilayah kepulauan seribu mempunyai perubahan peringkat dari hasil *cluster*ing dengan metode *K-Means* dan DBSCAN. Khusus untuk wilayah Kepulauan Seribu memang berada pada peringkat terakhir baik dengan menggunkaan metode *K-Means* ataupun dengan menggunakan metode DBSCAN. Hal ini bisa jadi disebabkan daerah Kepulauan Seribu yang tidak berbatasan dengan wilayah DKI Jakarta lainnya dengan kata lain wilayah ini dipisahkan oleh lautan yang cukup luas sehingga memungkinkan daerah ini mempunyai tingkat isolasi yang tinggi dibanding wilayah DKI Jakarta lainnya.

Untuk mengetahui seberapa baik kualitas dari *clustering* yang dihasilkan dari kedua metode *clustering* diatas maka dilakukan evaluasi hasil *cluster*ing yang menggunakan metode *Davies-Bouldin Index*. Dalam melakukan evaluasi hasil *clustering* menggunakan metode ini langkah yang dilakukan adalah menghitung nilai *Sum of Square Within Cluster* (SSW) kemudian Sum of Square Between *Cluster* (SSB) dan Ratio.

Nilai dari *Sum of Square Within Cluster* (SSW) digunakan untuk mengetahui kohesi dalam sebuah *cluster* ke -*i*. Kohesi didefinisikan sebagai sebuah jumlah dari kedekatan data terhadap titik pusat *cluster* dari sebuah *cluster* yang di ikutinya. Rumus yang digunakan untuk mendapatkan nilai SSW adalah sebagai berikut.

$$SSW_i = \frac{1}{m_i} \sum_{i=i}^{m_i} d(x_j, c_i)$$

Hasil dari perhitungan nilai SSW kedua metode tersebut terlihat pada tabel dibawah ini.

Tabel 11. Hasil SSW *K-Means*.

Hasil Perhitungan SSW		
Cluster	K-Means	DBSCAN
C0	0,002508634	0,210550092
C1	0,026986139	0,598439154
C2	0,004465785	0,373438126
C3	0,002637722	0,294674697

Tabel 11 memperlihatkan Nilai dari *Sum of Square Beetwen Cluster* (SSB) digunkan untuk mengetahui separasi antar *cluster*. Rumus untuk menghitung nilai SSB adalah sebagai berikut.

$$SSB_{i,j} = d(c_i, c_j)$$

Hasil dari perhitungan nilai SSB kedua metode tersebut terlihat pada tabel 12 dibawah ini.

Tabel 12. Hasil SSB *K-Means*.

Hasil SSB K-Means				
Cluster	C0	C1	C2	C3
C0	0	1.091040176	0.585842907	0.359554156
C1	1.091040176	0	0.601122164	0.696274215
C2	0.585842907	0.601122164	0	0.327134814
C3	0.359554156	0.696274215	0.327134814	0

Tabel 13. Hasil SSB DBSCAN.

Hasil SSB DBSCAN				
Cluster	C0	C1	C2	C3
C0	0	1.219	0.593007674	0.391939435
C1	1.218960309	0	0.642595501	0.743759309
C2	0.593007674	0.643	0	0.314930439
C3	0.391939435	0.744	0.314930439	0

Tabel 13 memperlihatkan hasil dari perhitungan *Ratio* (Rasio) yang digunakan untuk mengetahui nilai perbandingan antar *cluster* ke -i dan *cluster* ke -j. Rumus untuk menghitung nilai ratio yang dimiliki oleh masing-masing *cluster* adalah sebagai berikut.

$$R\ i,\ j = \frac{1}{mi}\ ssw_i$$

Hasil dari perhitungan nilai *Ratio* kedua metode tersebut terlihat pada tabel 14 dibawah ini.

Tabel 14. Hasil *Ratio K-Means*.

Ratio K-Means	
Ratio	Hasil
R0,1	0,101787948
R0,2	0,062471149
R0,3	0,101787948
R1,2	0,060883264
R1,3	0,052563026
R2,3	0,111875221

Tabel 15. Hasil Ratio DBSCAN

Ratio DBSCAN	
Ratio	Hasil
R0,1	1,211772079
R0,2	2,490865011
R0,3	3,768699789
R1,2	2,298649874
R1,3	1,98599473
R2,3	4,690248659

Tabel 15 merupakan Nilai *ratio* yang telah di dapatkan akan digunkan untuk mencari nilai *Davies Bouldin Index* (DBI) dengan menggunakan rumus di bawah ini.

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max\ (R_{i,j})$$

Hasil dari perhitungan nilai DBI tersebut akan digunakan sebagai perbandingan efektifitas *clustering* kedua metode yang digunakan yaitu *K-Means* dan DBSCAN. Semakin kecil nilai *Davies Bouldin Index* (DBI) yang diperoleh (Non Negatif >= 0), maka akan semakin baik *cluster* yang dihasilkan dari pengelompokan menggunakan kedua metode tersebut.

Hasil dari perhitungan nilai DBI kedua metode tersebut terlihat pada tabel 16 dibawah ini.

Tabel 16. Hasil DBI *K-Means* dan DBSCAN.

Metode	DBI
K-Means	0,027968805
DBSCAN	1,172562165

Maka perbandingan dari kedua hasil DBI tersebut diatas menunjukan bahwa Algoritma *K-Means* lebih baik dibanding algoritma DBSCAN dalam kasus ini dengan perbandingan 0,027968805 < 1,172562165.

4. Kesimpulan

Setelah melakukan perhitungan dan pengolahan data dengan menggunakan metode *K-Means* dan DBSCAN dalam peng*cluster*an data *covid-19* di wilayah DKI Jakarta dapat disimpulkan 1) Pola penyebaran virus *covid-19* di DKI Jakarta tidak hanya berdasarkan kepada luas daerah dari wilayah-wilayah yang ada di kota administrasi DKI Jakarta. 2. *Clustering* dengan metode *K-Means* menghasilkan 4 *cluster* dengan komposisi *cluster* 0 dengan jumlah anggota sebanyak 91, *cluster* 1 jumlah anggota sebanyak 23, *cluster* 2 jumlah anggota sebanyak 62, *cluster* 3 jumlah anggota sebanyak 91. Validasi nilai DBI sebesar 0.027968805. *Clustering* dengan metode DBSCAN menghasilkan *cluster* dengan komposisi *cluster* 0 dengan jumlah anggota sebanyak 87, *cluster* 1 dengan jumlah anggota sebanyak 23 *cluster* 2 dengan jumlah anggota sebanyak 63, *cluster* 3 dengan jumlah anggota sebanyak 94. Validasi nilai DBI sebesar 1,172562165. 3) Proses validasi hasil *cluster* menggunakan metode *Davies Bouldin Index* (DBI) menghasilkan nilai 0.027968805 untuk *K-Means* dan 1,172562165 untuk DBSCAN, maka nilai DBI yang terkecil terdapat pada metode *K-Means* yaitu 0.027968805 dibanding dengan metode DBSCAN dengan nilai 1,172562165. Dengan demikian dalam penelitian ini metode yang optimal dalam pembentukan *cluster* dari data *covid-19* di wilayah DKI Jakarta adalah metode *K-Means.*

Referensi

Anggara, M., Sujiani, H., & Nasution, H. (2016). Pemilihan Distance Measure Pada K-Means *Cluster*ing Untuk Pengelompokkan Member Di Alvaro Fitness. *JUSTIN (Jurnal Sistem Dan Teknologi Informasi)*, *1*(1), 1–6.

Aritonang, A. R., Sutarman, & Sihombing, P. (2015). Analisis Subspace *Cluster*ing Menggunakan DBSCAN dan SUBCLU Untuk Proyeksi Pekerjaan Alumni Perguruan Tinggi. *Teknovasi*, *02*, 33–60.

Bastian, A., Sujadi, H., & Febrianto, G. (2018). Penerapan Algoritma K-Means *Cluster*ing Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka). In *jsi.cs.ui.ac.id*. Retrieved February 24, 2021, from http://jsi.cs.ui.ac.id/index.php/jsi/article/view/566

Budiman, S., Safitri, D., & Ispriyanti, D. (2016). Perbandingan Metode K-Means Dan Metode Dbscan Pada Pengelompokan Rumah Kost Mahasiswa Di Kelurahan Tembalang Semarang. *None*, *5*(4), 757–762.

Fajriah, R. I., Sutisna, H., & Simpony, B. K. (2019). Perbandingan Distance Space Manhattan Dengan Euclidean Pada K - Means *Cluster*ing Dalam Menentukan Promosi. *Informatika, Bina Sarana Bsi, Universitas*, *4*(1), 36–49.

Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk *Cluster*ing Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *5*(1), 17–24. https://doi.org/10.25077/teknosi.v5i1.2019.17-24

Handayani, R. T., Arradini, D., Darmayanti, A. T., Widiyanto, A., & Atmojo, J. T. (2020). Pandemi covid-19, respon imun tubuh, dan herd immunity. *Jurnal Ilmiah Stikes Kendal*, *10*(3), 373–380.

Pratiwi[1], M., Agustini[3], P., Jendral, J., Yani, H., & 12 Palembang, N. (2015). Penerapan K-Means *Cluster*ing Untuk Memprediksi Minat Nasabah Pada Pt. Asuransi Jiwa Bersama 1912 Bumiputera Prabumulih. In *if.binadarma.ac.id*. Retrieved February 24, 2021, from http://if.binadarma.ac.id/sipi/jurnal/Jurnal-Jurnal Skripsi Lhorend.pdf

Putu, N., Merliana, E., & Santoso, A. J. (2016). *Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means*. 978–979.

Rahman, A. T., Wiranto, & Rini, A. (2017). Coal Trade Data *Cluster*ing Using K-Means (Case Study Pt. Global Bangkit Utama). *ITSMART: Jurnal Teknologi Dan Informasi*, *6*(1), 24–31. https://jurnal.uns.ac.id/itsmart/article/download/11296/11108

Sari, B. N., & Primajaya, A. (2019). Penerapan *Cluster*ing Dbscan Untuk Pertanian Padi Di Kabupaten Karawang. *Jurnal Informatika Dan Komputer*,*4*(1), 28–34. www.mapcoordinates.net/en.