

BAB II

LANDASAN TEORI

2.1. Landasan Teori

2.1.1. Covid-19

Coronavirus merupakan bagian dari virus yang menyerang penyakit pada manusia dan hewan. Virus yang menyerang manusia akan berdampak pada infeksi saluran pernapasan, flu hingga penyakit yang lebih serius seperti *Middle East Respiratory Syndrome* (MERS) dan Sindrom Pernafasan Akut Berat/Severe Acute Respiratory Syndrome (SARS). Jenis baru dari *coronavirus* yang pertama kali ditemukan di Wuhan, Cina pada Desember 2019 yaitu *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-COV2), dan mengakibatkan adanya penyakit *Coronavirus Disease-2019* (Covid-19) [9].

2.1.2. Vaksin Covid-19

Vaksin merupakan salah satu cara yang dapat digunakan dalam mencegah penyakit dan menjaga kondisi tubuh. Vaksin atau imunisasi memiliki manfaat dari fungsi uni yang ada didalam tubuh untuk mempelajari dan melawan kuman yang mengakibatkan penyakit didalam tubuh manusia. Oleh karena itu, vaksin dapat membantu manusia dalam menjaga kekebalan tubuh untuk melindungi diri dari infeksi dan tidak memiliki efek samping yang berbahaya [8]. Penelitian ini menggunakan penyebaran data vaksin yang sudah diberikan kepada masyarakat luas di masing-masing wilayah propinsi DKI Jakarta sebagai bahan evaluasi dan memastikan apakah vaksin Covid-19 mempengaruhi hasil pembentukan klaster yang dibentuk

2.1.3. *Data Mining*

Data mining adalah model informasi untuk penyimpanan *database*.

Dalam *data mining* dapat disebutkan sebagai langkah-langkah proses untuk mencari informasi tambahan yang sebelumnya belum terdapat dalam *database* serta model yang berguna untuk menemukan model data yang menjadikannya informasi penting dari hasil pengenalan model yang bermanfaat serta menarik dari data yang ada didalam *database* [10].

2.1.4. Kerangka Kerja CRISP-DM

Kerangka kerja CRISP-DM (*Cross-Industry Standard Process Model for Data Mining*) ialah metode yang sering digunakan dalam proses data mining dengan memiliki 6 tahap yang terlihat pada Gambar 2.1 berikut.



Gambar 2.1 Tahapan CRISP-DM
Sumber: Yogasetya Suhandi [10]

Dari Gambar 2.2 dapat dijelaskan tahapan-tahapan CRISP-DM adalah sebagai berikut [11]:

1. *Business Understanding*

Business Understanding merupakan tahap yang dilakukan dengan mempelajari tujuan bisnis, kondisi bisnis yang ada di tempat penelitian terkait permasalahan yang dapat diselesaikan dengan proses data mining.

2. *Data Understanding*

Data understanding merupakan tahap persiapan data, pada tahap ini melakukan pengecekan pada data yang akan digunakan yaitu dengan melakukan pengumpulan data awal dan identifikasi kualitas data. Pada

tahap ini akan dilakukan proses pendeskripsian dari tiap fitur yang terdapat dalam data yang digunakan.

3. *Data Preparation*

Data *preparation* merupakan tahap mempersiapkan data yang telah dikumpulkan. Pada tahap ini dilakukan proses identifikasi data, pemilihan data, pembersihan dan transformasi data. Setelah data sudah siap dapat dilanjutkan pada tahap *modelling*.

4. *Modelling*

Modeling merupakan tahap menerapkan algoritma yang digunakan untuk menghasilkan pola data yang sesuai dengan data penelitian.

5. *Evaluation*

Evaluation merupakan tahap untuk mengevaluasi model yang telah dibuat sehingga mengetahui model yang digunakan tepat atau tidak digunakan untuk proses data mining tertentu.

6. *Deployment*

Deployment atau penyebaran merupakan proses menggunakan model yang dihasilkan sebelumnya. Dalam tahap ini terdapat 2 jenis kegiatan penyebaran yaitu melakukan perencanaan dan pemantauan dari penyebaran hasil yang dilakukan. Jenis kegiatan selanjutnya menyelesaikan tugas penutup dengan membuat laporan akhir dan melakukan tinjauan proyek. Kedua kegiatan ini dapat dilakukan semua atau menyelesaikan salah satunya.

2.1.5. **Klustering**

Klustering merupakan teknik penggalian data dalam *data mining*. Fungsi dari klustering adalah untuk mengelompokkan data berdasarkan tingkat kemiripan dari data yang dapat pada kelompok tertentu dan meminimalkan kemiripan pada kelompok lain. Klustering dapat diartikan sebagai proses untuk melakukan pengamatan, mengelompokkan *record* dan kelas objek yang terdapat kemiripan.

[12]

2.1.6. Metode *K-Means*

K-Means merupakan bagian dari metode Klastering *non-hirarki* dan dimanfaatkan untuk pengelompokan data ke satu atau beberapa klaster. Penerapan *K-Means Klastering* yaitu ketika data yang terdapat ciri yang sama akan diklasterisasi ke sebuah klaster, sedangkan data yang memiliki ciri berbeda dikelompokkan dengan klaster lain. Hal tersebut dilakukan agar data pada satu klaster mempunyai variasi yang kecil. Berikut adalah langkah melakukan klastering menggunakan metode ini [12]:

1. Penentuan jumlah klaster (k) yang digunakan pada dataset
2. Penentuan pusat klaster (*centroid*)

Pada proses *centroid* ditentukan secara acak menggunakan formulasi untuk penentuan tujuan awal pada metode *K-Means Klastering*.

$$\frac{\text{Jumlah data}}{\text{Jumlah class}+1} \quad (2.1)$$

Keterangan:

Jumlah data adalah keseluruhan dari data yang dipakai.

Jumlah *class* adalah jumlah kelompok yang sebelumnya telah ditentukan.

Pada fase iterasi dipakai rumus rata-rata:

$$\bar{y}_j = \frac{1}{N_i} \sum_{k=0}^n x_{kj} \quad (2.2)$$

Keterangan:

\bar{y}_j = *centroid* ke- i pada variabel ke- j

N_i = jumlah data yang menjadi anggota klaster ke- i

i, k = klaster

j = variabel

x_{kj} = nilai data ke- k yang ada di dalam klaster tersebut untuk variabel ke- j

3. Pada setiap *record*, lakukan perhitungan jarak terdekat dengan *centroid*:

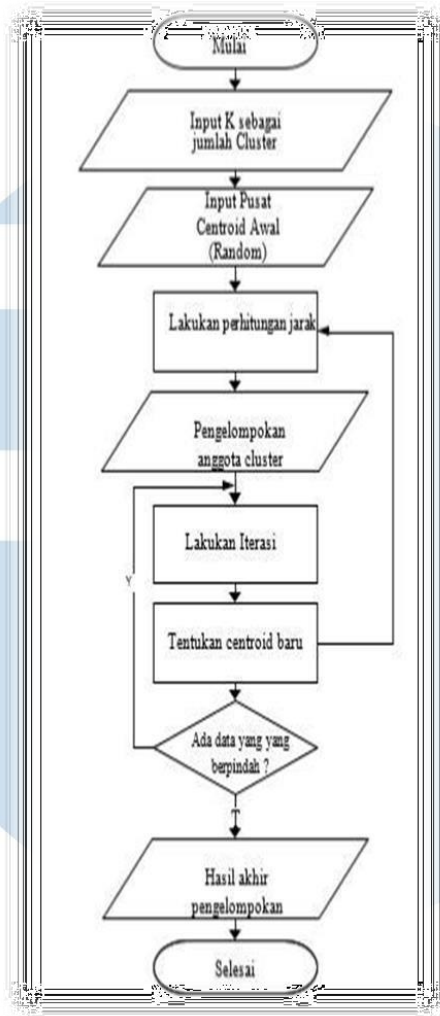
$$De = \sqrt{(xi - si)^2 + (yi - ti)^2} \quad (2.3)$$

Keterangan:

- De = *Euclidean Distance*
 i = Jumlah objek yang digunakan
 (x, y) = koordinat pada objek
 (s, t) = koordinat pada centroid

4. Lakukan pengelompokkan objek.
5. Lakukan iterasi langkah ke-2 hingga centroid bernilai optimal.
6. Selanjutnya lakukan perhitungan jarak pada setiap data ke setiap *centroid* dengan memakai rumus *euclidean distance*. Jika telah diketahui nilai jarak setiap data, maka kelompokkan data berdasarkan jarak terdekat dari centroid. Jika masih ada data yang belum masuk ke dalam kelompok klaster, maka ulangi dari menentukan centroid hingga tidak ada lagi data yang tidak masuk ke dalam kelompok klaster.
7. *K-means* hasil segmentasi data tidak bisa memberikan pola kelompok yang mewakili ciri bentuk alami data.
8. *K-means* dapat mengalami masalah ketika mengklasterkan data yang didalamnya mencakup *outlier*

Flowchart algoritma *K-Means* ditampilkan pada Gambar 2.2. yang memiliki langkah pertama yaitu melakukan *input* nilai K sebagai jumlah klaster yang akan digunakan. Kemudian, melakukan *input* pusat *centroid* awal. Setelah menentukan pusat *centroid* awal dilakukan perhitungan jarak. Hasil tersebut menghasilkan pengelompokan anggota klaster. Tahap selanjutnya melakukan iterasi sejumlah nilai K dan menentukan *centroid* baru. Berikutnya menghitung nilai jarak pada *centroid* baru. Jika terdapat data yang berpindah, maka akan berpindah pada proses iterasi. Jika tidak didapatkan hasil akhir pengelompokan.



Gambar 2.2 Langkah Kerja Algoritma K-Means

Sumber: Sulistyowati [12]

Dengan menggunakan *K-Means* dapat mempermudah pihak peneliti dalam menganalisis dan mengelompokan data, karena Algoritma *K-means* merupakan algoritma iteratif yang mencoba mempartisi kumpulan data ke dalam subkelompok (klaster) berbeda yang tidak tumpang tindih yang telah ditentukan sebelumnya di mana setiap titik data hanya dimiliki oleh satu kelompok [13]. Oleh karena itu, algoritma *K-Means* digunakan untuk mengetahui penyebaran kasus Covid-19 dengan melihat data nilai Covid-19 tahun 2020-2022

2.1.6.1. Metode Elbow

Merupakan cara dalam mendapatkan informasi dalam penentuan total klaster terbaik dengan persentase hasil komparasi antara jumlah klaster yang membentuk siku di sebuah titik yang akan digunakan pada klasterisasi *K-Means*. Selanjutnya menambahkan nilai klaster untuk digunakan sebagai model data untuk menentukan klaster terbaik. Dalam menentukan nilai klaster terbaik dilihat dari nilai SSE. SSE (*Sum of Square Error*) merupakan perhitungan yang dilakukan dengan menjumlahkan seluruh jarak dari setiap data dengan titik pusat klaster. Data yang seragam dapat dilihat dari nilai SSE yang diperoleh, jika nilai SSE kecil maka menghasilkan klaster yang baik [14]. SSE digunakan pada metode Elbow untuk memaksimalkan kinerja dari algoritma *K-means* [14]. Rumus SSE sesuai dengan Persamaan (2.4)

$$SSE = K = \sum_{k=1}^K \sum_{xi} |xi - ck|^2 \quad (2.4)$$

Ket:

K = Klaster ke-c

Xi = jarak data obyek ke-i

Ck= pusat klaster ke-i

Hasil perhitungan akan digunakan sebagai nilai pembanding antara jumlah klaster yang ditambahkan. Hasil dari tiap nilai klaster dapat ditampilkan dalam bentuk grafik. Klaster terbaik dapat dilihat dari nilai klaster pertama dan kedua yaitu jika nilai tersebut mengalami penurunan paling besar [15].

2.1.7. Metode Fuzzy C-Means

Fuzzy C-Means disebut sebagai suatu metode pengklasifikasian data yang seluruh titik data pada kelompok ditentukan melalui derajat keanggotaan. *Fuzzy c-means* merupakan algoritma pengklasteran yang terkontrol.

Teorema FCM merupakan salah satu metode yang dikenalkan oleh Jim Bezdek pada tahun 1981, metode FCM merupakan teknik dalam mengelompokkan data yang keberadaan data didalam kluster ditentukan oleh nilai keanggotaan dan nilai derajat keanggotaan dengan jangkauan nilai 0 sampai 1 [16]. Teorema FCM memiliki persamaan 2.5 sampai 2.9.

$$\sum_{j=1}^k \mu_{ij} \quad (2.5)$$

Dimana:

k = jumlah kluster

μ_{ij} = derajat keanggotaan

i = jumlah data

j = jumlah kluster

$$0 < \sum_{i=1}^n u_{ij} < n \quad (2.6)$$

Dimana:

μ_{ij} = derajat keanggotaan

n = jumlah data

$$u_{ij} = \frac{D(x, c_j)^{-2}}{\sum_{l=1}^k D(x, c_l)^{-2}} \quad (2.7)$$

Dimana:

c_j = Centroid kluster

D = Jarak antar data

w = parameter bobot pangkat

$$c_{ij} = \frac{\sum_{i=1}^N (u_{ij})^w x_{ij}}{\sum_{i=1}^N (u_{ij})^w} \quad (2.8)$$

$$\sum_{i=1}^N (u_i)^w$$

Dimana:

N = Jumlah data

$$J = \sum_{i=1}^N \sum_{l=1}^k (u_{ij})^w D(x_i, c_l)^2 \quad (2.9)$$

Dimana:

J : Jumlah dari kluster D : Jarak

N : Jumlah data

w : Bobot

2.1.8. *Davies-Bouldin Index (DBI)*

Indeks validitas Davies-Bouldin (DB) menghitung rata-rata nilai setiap titik pada himpunan data. Perhitungan nilai setiap titik adalah jumlah nilai compactness yang dibagi dengan jarak antara kedua titik pusat kluster sebagai separation. Jumlah kluster terbaik ditunjukkan dengan nilai DB yang semakin kecil [17].

2.1.9. *Calinski-Harabasz (CH)*

Indeks validitas *Calinski-Harabasz* (CH) menghitung perbandingan antara nilai *Sum of Square between cluster* (SSB) sebagai separation dan nilai *Sum of Square within-cluster* (SSW) sebagai *compactness* yang dikalikan dengan faktor normalisasi, yaitu selisih jumlah data dengan jumlah kluster dibagi dengan jumlah kluster dikurang satu. Jumlah kluster terbaik ditunjukkan dengan semakin besar nilai CH [17].

2.1.10. *Silhouette Index*

Secara umum, *Silhouette Index* atau indeks validitas *Silhouette* menghitung rata-rata nilai setiap titik pada himpunan data. Dengan lebih spesifik, perhitungan nilai setiap titik adalah selisih nilai *separation* dan *compactness* yang dibagi dengan maksimum antara

keduanya. Jumlah kluster yang terbaik ditunjukkan dengan nilai *Silhouette* yang semakin mendekati 1 [17].

2.1.11. Pemrograman Python

Python merupakan *high level language* yang sudah menjadi standar dalam bidang komputasi ilmiah. Python dapat digunakan di segala *operating system* dan dikatakan sebagai bahasa yang fleksibel dan mudah dipelajari. Python lebih mudah dibaca jika dibandingkan dengan bahasa pemrograman lain [18].

Dalam *Python* terdapat *library*, *library* merupakan kumpulan modul yang berisi kumpulan kode yang dapat digunakan berulang kali dalam program yang berbeda [19]. Dalam *library* pada Python yang tersedia diantaranya yaitu numPy, Pandas, matplotlib, sklearn, serta menyediakan berbagai algoritma dan menghasilkan *output* berupa visualisasi dalam bentuk grafis dan table [19].

2.1.11.1 Numpy Library

Numpy merupakan *library* yang digunakan dalam melakukan perhitungan saintik seperti matriks. [19]

2.1.11.2 Pandas Library

Pandas merupakan *library* yang digunakan dalam membaca data dengan berbagai jenis format seperti .csv, .txt dan sebagainya. [19]

2.1.11.3 Matplotlib Library

Matplotlib merupakan *library* yang digunakan dalam membuat grafik plot dari hasil visualisasi. [19]

2.1.11.4 Seaborn Library

Seaborn merupakan *library* yang digunakan dalam membuat melakukan visualisasi data berupa garfik. [19]

2.1.11.5 Sklearn Library

Sklearn merupakan *library* yang menyediakan berbagai metode dan algoritma untuk membuat *machine learning* [19].

2.1.12. *Pearson Correlation*

Untuk mencari dan menemukan indikasi penyebab peningkatan jumlah pasien positif covid maka pada penelitian dilanjutkan dengan melakukan pembuktiannya melalui *pearson correlation* untuk mengetahui apakah jumlah pasien yang belum melakukan vaksinasi dapat menyebabkan peningkatan jumlah pasien pada klaster merah.

Pearson Correlation atau Korelasi Pearson adalah ukuran yang dapat menentukan signifikansi dari hubungan antara dua variabel yang diukur. Korelasi Pearson digunakan untuk mengetahui apakah variabel yang digunakan memiliki korelasi dengan variabel lain. Dalam penelitian ini Korelasi Pearson untuk menentukan apakah ada korelasi antara data masyarakat yang belum melakukan vaksinasi dengan masyarakat yang terindikasi kasus positif covid, dengan itu dapat menghitung koefisien korelasinya. Metodologi untuk studi korelasi:

- *Parametric Correlation*: pengujian yang menilai ketergantungan linier antara dua variabel (x dan y) dan ketergantungan pada distribusi data dikenal sebagai uji korelasi parametrik.
- *Non-Parametric Correlation*: parametrik ini mengacu pada Kendall(tau) dan Spearman(rho), yang merupakan koefisien korelasi berbasis peringkat.

Formula dari *Pearson Correlation*:

$$r = \frac{\sum(x-m_x)(y-m_y)}{\sqrt{\sum(x-m_x)^2 \sum(y-m_y)^2}} \quad (2.10)$$

Dimana:

r = nilai koefisien korelasi [20]

Berdasarkan koefisien korelasi pearson terdapat kriteria yang digunakan untuk menentukan derajat kuat atau tidaknya hubungan antara dua variabel. Aturan sederhana atau yang biasa disebut dengan *rule of thumb* untuk kriteria penentuan derajat kuat atau tidaknya hubungan antara dua variabel:

$0 < |r| < 0,490$ mempunyai hubungan lemah

$0,50 < |r| < 0,790$ mempunyai hubungan sedang

$0,80 < |r| < 1$ mempunyai hubungan kuat

Dengan beberapa catatan tambahan:

- r terletak di antara -1 (korelasi negative) dan 1 (korelasi positif).
- $r = 0$ yang berarti tidak ada korelasi
- Tidak dapat diterapkan pada variabel ordinal [21]

2.1.8.1 Heatmap

Untuk memvisualisasikan hasil *Pearson Correlation* salah satunya dapat menggunakan *Heatmap*. *Heatmap* adalah *tools* yang digunakan untuk membuat visual grafik secara menarik yang dapat menawarkan wawasan tentang pola data dan memungkinkan untuk dengan mudah menemukan *outlier* dalam himpunan data.

Heatmap matriks korelasi memungkinkan kita untuk menentukan seberapa baik atau buruk fitur dalam himpunan data berkorelasi satu sama lain dan apakah korelasinya positif atau negatif. Di dalam *Python* menawarkan banyak alat visualisasi data, tetapi *library Seaborn* adalah salah satu yang paling populer dan ramah pengguna. Dengan panggilan fungsi tunggal dan himpunan data, memudahkan untuk menghasilkan grafik *heatmap* [22].

2.2. Penelitian Terdahulu

Berikut adalah tabel perbandingan penelitian terdahulu:

Tabel 2.1 Penelitian Terdahulu

No	Judul Penelitian	Jurnal	Tahun	Penulis	Hasil Penelitian
1.	Analisis Data Mining Untuk Klustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means	<ul style="list-style-type: none"> • Jurnal: Jurnal Teknologi Dan Sistem Informasi • Volume: 2 • Halaman: 100 – 108 	2021	Z. Nabila, A. R. Isnain, P. Permata dan Z. Abidin,	Hasil penelitian adalah pada perhitungan manual dan bantuan <i>tools</i> Rapid Miner mendapatkan nilai mendekati nilai DBI0 yang berarti kluster yang dievaluasi menghasilkan kluster terbaik. [7].

2.	Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means	<ul style="list-style-type: none"> • Jurnal: <i>Journal of Information Technology and Computer Science (INTECOMS)</i> • Volume: 1, Hal : 72-77 	2018	R. R.Putra dan Cendra Wadisman	Hasil penelitian berupa nilai <i>centroid</i> terbesar sebagai pelanggan potensial berdasarkan algoritma <i>K-Means</i> [22].
3.	Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah	<ul style="list-style-type: none"> • Jurnal: <i>IT Journal Research and Development</i> • Volume: 3 • Halaman: 22 – 31 	2018	R.Rosmini, A.Fadlil dan S.Sunardi	Hasil penelitiannya yaitu 2 <i>cluster</i> yang terdiri dari <i>cluster A</i> dan <i>cluster B</i> yang masing-masing memiliki jumlah 5 mahasiswa yang lulus tepat waktu dan tidak lulus tepat waktu [23].
4.	<i>Implementation of K-Means Algorithm for Klustering Corn Planting Feasibility Area in South Lampung Regency</i>	<ul style="list-style-type: none"> • Jurnal: <i>Journal of Physics: Conference Series</i> • Volume: 1751 • Halaman: 1-9 	2021	A. Aldino, D. Darwis, A. Prastowo dan C. Sujana	Hasil penelitian ini berupa wilayah dengan jumlah panen terbanyak wilayah dengan jumlah panen terendah adalah Candipuro [24].
5.	<i>Research on K-Value Selection Method of K-Means Klustering Algorithm</i>	<ul style="list-style-type: none"> • Jurnal: <i>J Multidisciplinary Scientific Journal</i> • Volume: 2 • Halaman: 226-235 	2019	C. Yuan dan H. Yang	Dari hasil penelitian algoritma <i>K-Means Klustering</i> yang dievaluasi mendapatkan waktu eksekusi tercepat pada metode <i>Elbow</i> dibandingkan dengan <i>GapStatistic</i> , <i>SilhouetteCoefficient</i> dan <i>Canopy</i> [25].
6.	<i>Declarative Knowledge Discovery in Databases via Meta-Learning – Towards Advanced Analytics</i>	<ul style="list-style-type: none"> • Jurnal: <i>Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)</i> • Volume: - • Halaman: 1-12 	2021	Dietmar Seipel, dan Martin Atzumuller.	Hasil dari penelitian ini berupa penggunaan kernel <i>vikami ne</i> mendapatkan <i>Sampras</i> sebesar 85% dari semua kasus dan nilai <i>Agassi</i> 0 [26].

UNIVERSITAS
MULTIMEDIA
NUSANTARA

7.	Klasterisasi Persebaran Virus Corona(Covid-19) DiDKI Jakarta Menggunakan MetodeK-Means	<ul style="list-style-type: none"> • Jurnal: <i>Fountain of Informatics Journal</i> • Volume: 5 • Halaman: 152-59 	2020	Achmad Solichin, Khansa Khairunnisa	Penerapan metode K-means untuk melakukan klasterisasi kasus Covid-19 di DKI Jakarta terdapat 9 klaster yang direkomendasikan dalam mengambil keputusan strategis dalam mengurangi persebaran virus Corona di DKI Jakarta.
8.	Komparasi Algoritma K-Means dan K-Medoids Untuk Pengelompokan Penyebaran Covid-19 di Indonesia	<ul style="list-style-type: none"> • Jurnal: <i>Journal of Applied Informatics and Computing (JAIC)</i> • Volume: 5 • Halaman: 133-138 	2021	Anisa Fira, Chaerur Rozikin, Garno	Hasil penelitian yaitu memiliki klaster optimal sebanyak 2 klaster, dengan menggunakan algoritma K-Means dimana klaster 1 beranggotakan 2 wilayah dan dikategorikan tinggi. Perbandingan metode tersebut menghasilkan nilai Silhouette Coefficient 0,207 pada metode K- means dan 0,347 pada K-Medoids. [6]

Merujuk kepada daftar penelitian terdahulu pada tabel 2.1 Penelitian Terdahulu, penelitian ini memberikan hasil sebagai berikut:

1. Penelitian ini megadaptasi penerapan data mining model klaterisasi menggunakan algoritma K-Means dengan pengujian menggunakan metode elbow merujuk kepada penelitian "Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means" oleh Achmad Solichin, dan Khansa Khairunnisa dengan menghasilkan 9 klaster yang direkomendasikan dalam mengambil keputusan strategis untuk mengurangi persebaran virus Corona di DKI Jakarta.
2. Penelitian ini megadaptasi penerapan data mining model klaterisasi dengan membandingkan algoritma K-Means dan K-Medoids dengan pengujian menggunakan Silhouette Coefficient, merujuk kepada penelitian "Komparasi Algoritma K-Means dan K-Medoids Untuk Pengelompokan Penyebaran Covid-19 di Indonesia" oleh Anisa Fira, Chaerur Rozikin, dan Garno dengan pengujian klaster yang terbentuk menggunakan Silhouette Coefficient dan menghasilkan nilai Silhouette Coefficient sebesar 0,207 pada metode k- means

dan nilai 0,347 pada k-medoids.

3. Perbedaan dengan penelitian terdahulu yang disebutkan diatas, penelitian ini melakukan perbandingan 2 algoritma klastering lainnya yaitu K-Means dan Fuzzy C-Means untuk membentuk klaster penyebaran epidemi Covid-19 di wilayah DKI Jakarta dan untuk pengujian klasterisasi yang terbentuk menggunakan DBI (Davies-Bouldin Index) dan CH (Calinski-Harabasz) serta uji Silhouette.
4. Pada penelitian ini menghasilkan sebanyak 3 klaster yaitu klaster merah (daerah rawan), kuning (daerah cukup rawan) dan hijau (daerah aman) yang nantinya dapat digunakan sebagai referensi untuk menentukan kebijakan di masing-masing wilayah kelurahan dan kecamatan di DKI Jakarta untuk dapat mengurangi pertumbuhan penyebaran Covid-19.
5. Penelitian ini melakukan pencarian wawasan apakah penyebab peningkatan jumlah pasien positif Covid-19 dapat disebabkan oleh belum dilakukannya vaksinasi pada penduduk di wilayah DKI Jakarta, dan pengujian hasil yang didapatkan dilakukan menggunakan metode Pearson Correlation untuk mengetahui apakah jumlah pasien yang belum melakukan vaksinasi dapat menyebabkan peningkatan jumlah pasien pada klaster merah.
6. Hal lain dari penelitian ini menghasilkan wawasan hubungan antara populasi di masing-masing wilayah daerah DKI Jakarta apakah berpengaruh terhadap pertumbuhan tingkat kasus positif Covid-19.