

PENERAPAN ALGORITMA K-MEANS PADA CLUSTERING VAKSINASI COVID-19 DAERAH JAWA TIMUR

Michael Sitorus ¹, Cornelia Antonieta.DC. ², Cyntia Larasati ³

^{1,2,3} Institut Teknologi dan Bisnis Bank Rakyat Indonesia, Jakarta, Indonesia

Email: michael.sitorus@bri-institute.ac.id ¹,

corneliaantonietadacosta15@gmail.com ², clarasati099@gmail.com ³

Abstract

Entering the era of the Covid-19 pandemic, the government has intensively implemented a vaccination program to date. The Covid-19 vaccination program is carried out as an effort to boost the immune system, reduce the risk of transmission, reduce the severe impact of the virus, to achieve group immunity. In its own implementation, the Covid-19 vaccination is regulated by the regional government in each province with a policy that requires the Covid-19 vaccination to be vaccinated twice for everyone who meets certain criteria. This study aims to cluster the implementation of vaccination in all areas of East Java province in 2021. The method used in conducting this clustering is the K-Means algorithm. From the results of the study, the results of the division or clustering of regions into three clusters were C1 for the area with the lowest vaccination, namely Pasuruan Regency, C2 for the area with moderate vaccination, namely Kediri City, and C3 for the highest vaccination area, namely Surabaya City. The clustering results obtained based on the K-Means algorithm can be used as input for the East Java Provincial government in evaluating the implementation of the Covid-19 vaccination.

Keywords: Covid-19, Vaccination, Cluster, K-Means

Abstrak

Memasuki era pandemi Covid-19, pemerintah gencar melaksanakan program vaksinasi hingga saat ini. Program vaksinasi Covid-19 ini dilakukan sebagai upaya untuk meningkatkan sistem kekebalan tubuh, mengurangi risiko penyaluran, mengurangi dampak berat dari virus, hingga mencapai kekebalan kelompok. Pada pelaksanaannya sendiri, vaksinasi Covid-19 ini diatur oleh pemerintah daerah di masing-masing provinsi dengan kebijakan yang mewajibkan untuk melakukan vaksinasi Covid-19 sebanyak dua kali bagi setiap orang yang memenuhi kriteria tertentu. Penelitian ini bertujuan untuk mengklusterisasi pelaksanaan vaksinasi di seluruh wilayah provinsi Jawa Timur pada tahun 2021. Metode yang digunakan dalam melakukan klastering ini adalah algoritma K-Means. Dari hasil penelitian diperoleh hasil pembagian atau klasterisasi daerah menjadi tiga klaster yaitu C1 untuk daerah dengan vaksinasi terendah yaitu Kabupaten Pasuruan, C2 untuk daerah dengan vaksinasi sedang yaitu Kota Kediri, dan C3 untuk daerah vaksinasi tertinggi yaitu Kota Surabaya. Hasil klasterisasi yang didapat berdasarkan algoritma K-Means ini, dapat digunakan untuk masukan bagi pemerintah Provinsi Jawa Timur dalam evaluasi pelaksanaan vaksinasi Covid-19.

Kata Kunci: Covid-19, Vaksinasi, Klaster, K-Means.

Penerapan Algoritma K-Means Pada Clustering Vaksinasi Covid-19 Daerah Jawa Timur

**Michael Sitorus,
Cornelia Antonieta.DC.,
Cyntia Larasati**

Jurnal Teknosains
Kodepena
pp. 22-30



1. PENDAHULUAN

Coronavirus adalah virus RNA dengan ukuran partikel 120- 160 nm. Virus ini utamanya menginfeksi hewan, termasuk di antaranya adalah kelelawar dan unta. Sebelum terjadinya wabah COVID-19, ada 6 jenis *coronavirus* yang dapat menginfeksi manusia, yaitu *alphacoronavirus 229E*, *alphacoronavirus NL63*, *betacoronavirus OC43*, *betacoronavirus HKU1*, *Severe Acute Respiratory Illness Coronavirus (SARS-CoV)*, dan *Middle East Respiratory Syndrome Coronavirus (MERS-CoV)* (Riedel S, 1954). Sejak kasus pertama di Wuhan, terjadi peningkatan kasus COVID-19 di China setiap hari dan memuncak diantara akhir Januari hingga awal Februari 2020. Awalnya kebanyakan laporan datang dari Hubei dan provinsi di sekitar, kemudian bertambah hingga ke provinsi- provinsi lain dan seluruh China (Wu Z, 2020).

Infeksi Covid-19 dapat menular melalui percikan air liur (*droplet*) yang dikeluarkan ketika orang yang terkonfirmasi Covid-19 mengalami bersin, batuk, maupun berbicara. Selain itu, percikan air liur dapat menempel pada benda sehingga, apabila seseorang menyentuh permukaan benda yang didalamnya terdapat partikel Covid-19 dari orang terkonfirmasi, maka dapat menyebabkan penularan Covid-19 yang lebih masif lagi. Transmisi Covid-19 yang sangat mudah ini lah yang kemudian menjadikan pertambahan kasus positif Covid-19 kian hari kian melonjak. Dengan adanya pandemi Covid-19 ini, pemerintah menetapkan kebijakan untuk membatasi segala kegiatan kemasyarakatan untuk mengurangi kemungkinan penularan Covid-19 pada masyarakat. Pemerintah juga menetapkan protokol kesehatan pencegahan Covid-19 dengan 3M, yaitu mengenakan masker, mencuci tangan dengan sabun ataupun *handsanitizer*, serta menjaga jarak aman sosialisasi minimal 1 meter. Kebijakan lain yang diterapkan oleh pemerintah untuk mengurangi lonjakan kasus Covid-19 adalah memberikan vaksin Covid-19 kepada seluruh warga Indonesia. Pemberian vaksin bertujuan untuk memunculkan respon kekebalan tubuh seseorang terhadap serangan virus *SARS-Covid-19* sehingga, tubuh dapat melawan infeksi virus Covid-19. Tentunya, sistem kekebalan tubuh terhadap Covid-19 setelah divaksin tidak serta merta dapat terbentuk secara instan, protokol kesehatan 3M yang dicanangkan pemerintah haruslah tetap dilaksanakan untuk memberikan perlindungan maksimal terhadap serangan Covid- 19 (KEMENKES, 2021).

Salah satu penerapan vaksinasi Covid-19 tersebut adalah di Provinsi Jawa Timur. Gubernur Jawa Timur Khofifah Indar Parawansa menyatakan bahwa cakupan program vaksinasi Covid-19 di Jawa Timur paling tinggi jika dibandingkan dengan cakupan vaksinasi provinsi-provinsi lain di Indonesia. Khofifah pun mengimbau warga yang belum vaksin untuk segera ikut program vaksinasi. Menurut data Satuan Tugas Penanganan Covid-19, jumlah warga yang sudah mendapat suntikan dosis pertama vaksin Covid-19 di Jawa Timur sebanyak 7.058.237 orang, lebih banyak dibandingkan di DKI Jakarta (6.693.688 orang), Jawa Barat (5.134.735 orang), Jawa Tengah (4.523.284 orang), dan Bali (2.946.919 orang). Warga yang sudah mendapat dua suntikan vaksin atau sudah selesai menjalani vaksinasi di Jawa Timur tercatat sebanyak 2.694.731 orang, lebih banyak dibandingkan di Jawa Barat (2.315.419 orang), Jawa Tengah (2.311.019 orang), DKI Jakarta (2.041.918 orang), dan Bali (783.613 orang). Gubernur Jawa Timur optimistis dengan cakupan vaksinasi yang sudah dicapai Jawa Timur bisa mewujudkan kekebalan komunal terhadap Covid-19 pada Agustus 2021 (Teguh Firmansyah, 2021). Pada kasus ini dengan tersedianya data

mentah vaksinasi Covid-19 di provinsi Jawa Timur yang dapat diolah, penelitian ini dapat dikembangkan untuk membantu dalam melihat persebaran dan perkembangan vaksinasi Covid-19 di Provinsi Jawa Timur menggunakan metode *clustering* K-Means.

Data Mining

Data Mining adalah suatu proses pengerukan atau pengumpulan informasi penting dari suatu data yang besar. Proses data mining seringkali menggunakan metode statistika, matematika, hingga memanfaatkan Teknologi *Artificial Intelligence*. Nama alternatifnya yaitu *Knowledge discovery (mining) in databases (KDD)*, *knowledge extraction*, *data/pattern analysis*, *data archeology*, *data dredging*, *information harvesting*, *business intelligence*, dan lain-lain (Honakan, 2018).

Clustering

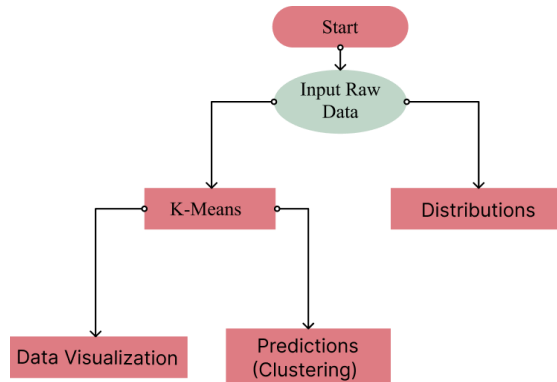
Clustering atau *clusterisasi* biasanya digunakan pada business intelligence, pengenalan pola citra, *web search*, bidang ilmu biologi, dan untuk keamanan (*security*). *Clustering* adalah proses pengelompokan data ke dalam beberapa *cluster* sehingga data-data di suatu *cluster* memiliki kemiripan maksimum. Cara kerja teknik ini ialah mengelompokkan sekumpulan data ke dalam kelas-kelas atau *cluster-cluster*, yang mana objek-objek yang ada pada kelas tersebut memiliki similaritas yang tinggi jika dibandingkan dengan objek lain yang ada dalam kelas tersebut, namun memiliki similaritas yang rendah jika dibandingkan dengan objek yang ada di kelas/*cluster* lain (N. Wiliani, 2019).

K-Means Clustering

K-Means Clustering adalah suatu metode penganalisaan data atau metode *Data Mining* yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Pada algoritma pembelajaran ini, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. Pembelajaran ini termasuk dalam *unsupervised learning*. Masukan yang diterima adalah data atau objek dan k buah kelompok (*cluster*) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Pada setiap *cluster* terdapat titik pusat (*centroid*) yang merepresentasikan *cluster* tersebut. Algoritma ini menggunakan masukan berupa jumlah k dan sekumpulan data set dari sekian objek untuk dimasukkan ke dalam k kelas/*cluster* (L. Fadilla, 2011). Cara kerja *K-means* adalah sebagai berikut. Pertama algoritma ini akan memilih sejumlah k objek dari beberapa objek yang ada pada sekumpulan data. Masing-masing objek yang terpilih, merepresentasikan nilai rata-rata dari sejumlah k *cluster*. Kemudian objek yang tersisa, masing-masing akan ditugaskan ke dalam *cluster* yang sudah ditentukan berdasarkan similaritasnya dengan nilai rata-rata masing-masing *cluster*. Setelah itu, dilakukan perubahan nilai rata-rata *cluster* yang sudah menjadi beberapa anggota *cluster*. Ketiga langkah di atas dilakukan secara iteratif (berulang) hingga tidak ada lagi perubahan nilai rata-rata dan seluruh data sudah didistribusikan ke masing-masing *cluster* yang ada.

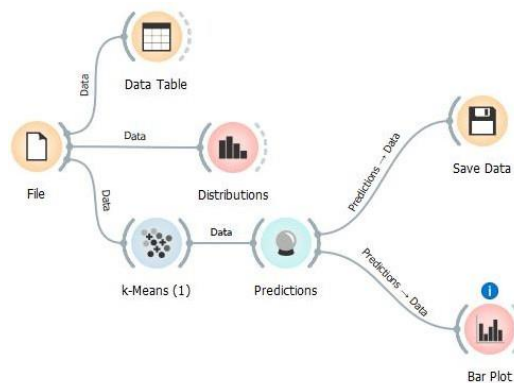
2. METODE PENELITIAN

Dalam penelitian ini, kami menggunakan metode penelitian kuantitatif. Metode penelitian kuantitatif adalah penelitian yang menghasilkan penemuan melalui prosedur-prosedur statistik atau pengukuran (Sujarweni, 2014). Pelaksanaan penelitian ini terdapat beberapa tahapan yang dilakukan yaitu mulai dari *input raw data*, *K-Means* yang terdiri dari *Bar Plot* dan *Predictions (Clustering)*, serta *Distributions* seperti pada gambar 1



Gambar 1. Tahapan Penelitian

Data penelitian ini berasal dari Portal Dinas Kesehatan Pemerintah Jawa Timur yang menyajikan data vaksinasi Covid-19 daerah Jawa Timur pada bulan Januari – Maret 2021. Adapun untuk proses yang dilakukan pada penelitian ini menggunakan aplikasi Orange dengan *workflow* seperti pada gambar 2 berikut.



Gambar 2. Workflow Proses *Clustering* Vaksinasi Covid-19 Daerah Jawa Timur

A. *Input Raw Data*

Data yang dikumpulkan merupakan data vaksinasi Covid- 19 daerah Jawa Timur pada bulan Januari – Maret. Dimana daalm data ini terdapat: (1) City or Regency, (2) Vaccinate, (3) Fully Vaccinated, (4) Total, (5) Date

B. *K-Means*

K-Means adalah algoritma yang digunakan untuk mengelompokkan beberapa data ke dalam kelompok- kelompok. Hal ini menjelaskan bahwa data pada satu kelompok memiliki karakteristik yang sama dan berbeda dengan

data pada kelompok lainnya (Bryan Orleans, 2022). Algoritma ini menggunakan proses iteratif untuk mendapatkan basis data *cluster*. Selain itu, algoritma *K-Means* membutuhkan jumlah *cluster* awal yang diinginkan untuk menjadi *input* dan menghasilkan titik *centroid* akhir sebagai *output*.

C. Bar Plot

Bar Plot atau diagram batang adalah grafik yang mewakili kategori data dengan batang persegi panjang dengan panjang dan tinggi yang sebanding dengan nilai yang diwakilinya. *Bar plot* dapat diplot secara *horizontal* atau *vertikal*. *Bar plot* ini menggambarkan perbandingan antara kategori diskrit. Salah satu sumbu plot mewakili kategori tertentu yang dibandingkan, sedangkan sumbu lainnya mewakili nilai terukur yang sesuai dengan kategori tersebut (Nikhila, 2022).

D. Clustering

Clustering merupakan proses pengelompokan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum (P. N. Tan, 2006). *Clustering* merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan *cluster*, dimana objek tersebut memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan *cluster* yang lain. Proses *clustering* ini digunakan untuk membantu memprediksi dan menganalisa keberhasilan vaksinasi Covid-19 daerah Jawa Timur pada bulan Januari – Maret 2021.

E. Distributions

Distribusi frekuensi adalah uraian berupa suatu tabel yang terdiri dari banyaknya kejadian frekuensi (*cases*) didistribusikan ke dalam kelompok-kelompok (kelas-kelas) yang berbeda (Budiyuwono, 1996). Data tersebut didapatkan dari hasil penelitian dan diolah sesuai dengan kebutuhan.

3. HASIL DAN PEMBAHASAN

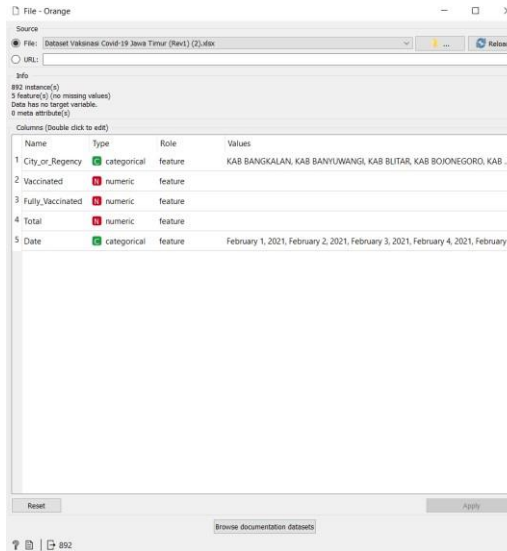
Dengan menggunakan metodologi yang telah dipaparkan pada bab sebelumnya, diperoleh hasil analisis dan perhitungan sebagai berikut :

A. Input Raw Data

Berdasarkan data pada gambar 3, dapat dilihat bahwa *raw data* vaksinasi Covid-19 di Jawa Timur memiliki lima atribut yakni :

1. *City_or_Regency* yang berisi kota atau kabupaten di Jawa Timur dengan tipe data *categorical*
2. *Vaccinated* yang berisi jumlah masyarakat yang sudah menerima vaksinasi Covid-19 di Provinsi Jawa Timur dengan tipe data *numeric*
3. *Fully-vaccinated* yang berisi jumlah masyarakat yang sudah menerima vaksinasi Covid-19 secara penuh di Provinsi Jawa Timur dengan tipe data *numeric*
4. *Total* yang berisi total masyarakat yang sudah divaksin Covid-19 baik secara penuh maupun masih parsial di di Provinsi Jawa Timur dengan tipe data *numeric*
5. *Date* yang berisi tanggal masuknya data vaksinasi Covid-19 di Jawa

Timur dengan tipe data *categorical*



Gambar 3. Input Raw Data pada Orange

B. Implementasi K-Means

1. Centroid Data

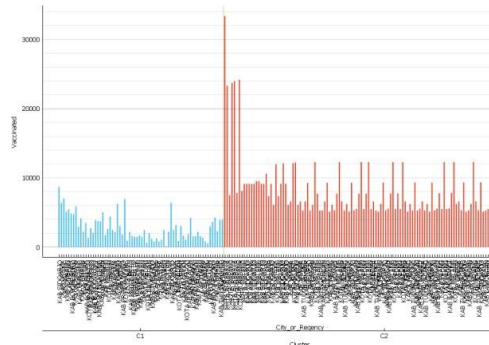
Dalam penerapan algoritma K-means dihasilkan nilai titik tengah atau centroid dari data yang didapat. Penentuan cluster dibagi atas tiga bagian yakni cluster rendah (C1), cluster sedang (C2) dan cluster tingkat tinggi (C3). Maka nilai titik tengah atau centroid juga terdapat 3 titik. Penentuan titik cluster ini dilakukan dengan mengambil nilai terkecil (minimum) untuk cluster rendah (C1), nilai rata-rata (average) untuk cluster sedang (C2) dan nilai terbesar (maksimum) untuk cluster tinggi (C3). Nilai titik tersebut dapat diketahui pada tabel berikut:

Tabel 1. Centroid Data (Iterasi 1)
Centroid

Min (C1)	4821	0,489175
Average (C2)	3073,959	0,551852
Max (C3)	38756	0,684341

2. Visualisasi Data

Pada data kali ini, visualisasi yang dilakukan adalah menggunakan grafik *Bar Plot* dengan nilai *values Vaccinated* dan *Group by Cluster* serta anotasi *none* dengan visualisasi sebagai berikut :



Gambar 4. Visualisasi data pada Orange

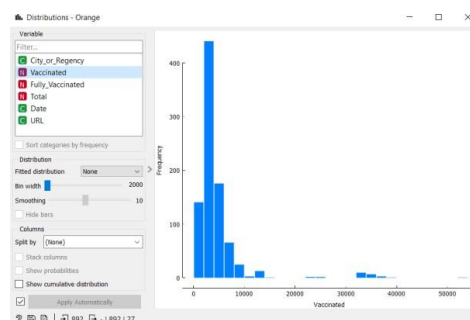
3. Clustering Data

Dengan menggunakan *centroid* tersebut maka dapat di *cluster* data yang telah didapat menjadi 3 *cluster*. Proses *cluster* dengan mengambil jarak terdekat dari setiap data yang diolah. Dari data jumlah kota/kabupaten yang telah melaksanakan vaksinasi Covid-19 pada Provinsi Jawa Timur didapatkan pengelompokan pada iterasi 1 untuk 3 *cluster* tersebut. *Cluster* yang memiliki nilai terendah dalam pelaksanaan vaksinasi Covid-19 di Jawa Timur (C1) adalah Kabupaten Pasuruan. *Cluster* yang memiliki nilai sedang (C2) yakni Kota Kediri dan *cluster* yang memiliki nilai tinggi (C3) adalah Kota Surabaya. Berikut tabel *clustering* data tersebut :

Proses *K-means* akan terus beriterasi sampai pengelompokan data sama dengan pengelompokan data iterasi sebelumnya. Dengan kata lain, proses akan terus melakukan iterasi sampai data pada iterasi terakhir sama dengan iterasi sebelumnya. Pada iterasi 1 diperoleh *cluster* data jumlah vaksinasi Covid-19 berdasarkan Kota/Kabupaten. Proses iterasi tersebut berhenti pada iterasi berikutnya, pada iterasi berikutnya akan dilakukan proses mencari nilai titik tengah atau *centroid*. Setelah mendapatkan nilai titik tengah atau *centroid*, proses sama dilakukan dengan mencari jarak terdekat. Proses pencarian jarak terpendek.

4. Distribusi Data

Pengelompokan data yang dilakukan terhadap 3 *cluster* dengan iterasi 1 didapatkan hasil yang sama dan distribusinya dapat dilihat pada grafik berikut :

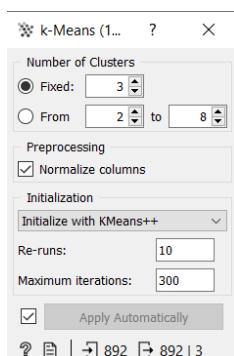


Gambar 5. Distribusi data pada Orange

5. Hasil Implementasi *Tool* Orange

Dalam pengelompokan/*clustering*, algoritma K- means digunakan untuk mengelompokan data berdasarkan pada atribut. Prinsip utama dari algoritma ini adalah menyusun k buah prototipe/pusat massa

(centroid)/rata-rata (mean) dari sekumpulan data berdimensi n . Aplikasi Orange akan membantu proses pengelompokan/clustering dan ditunjukkan pada berikut :



Gambar 6. Distribusi data pada Orange

Dari keseluruhan proses yang telah dipaparkan, diperoleh hasil bahwa daerah di Provinsi Jawa Timur yang berhasil menerapkan vaksinasi Covid-19 dengan keberhasilan tertinggi adalah Kota Surabaya, kemudian daerah yang masih dinilai sedang dalam menerapkan vaksinasi Covid-19 adalah Kota Kediri dan daerah dengan keberhasilan rendah adalah Kabupaten Pasuruan.

4. PENUTUP

Untuk melakukan pengelompokan keberhasilan pelaksanaan vaksinasi Covid-19 khususnya di Provinsi Jawa Timur dapat menerapkan metode *clustering K-means*. Data yang masih mentah diolah menggunakan *Ms. Excel* untuk ditentukan nilai *centroid* dalam 3 *cluster* yaitu *cluster* rendah (C1), *cluster* sedang (C2) dan *cluster* tinggi (C3). Hasil yang didapat dari penelitian dapat menjadi masukan kepada pemerintah, khususnya Kabupaten Pasuruan yang menjadi *cluster* terendah dalam pelaksanaan vaksinasi Covid-19 di Provinsi Jawa Timur agar segera melakukan tindakan lebih lanjut lagi.

5. DAFTAR PUSTAKA

Bryan Orleans, Edi Purnomo Putra. (2022). *Clustering Algoritma (K-Means)*. Retrieved 27 Juni 2022 from BINUS Website: <https://sis.binus.ac.id/2022/01/31/clustering-algoritma-k-means/>

Budiyuwono, N. (1996). *Pengantar Statistik Ekonomi & Perusahaan* (Edisi Pertama ed.). Yogyakarta. Retrieved 29 Juni 2022 from <https://localonsite.wordpress.com/2013/09/26/distribusi-frekuensi-dan-grafik-part-1/>.

GeeksforGeeks. (2020). *Bar Plot in Matplotlib*. Retrieved 29 Juni 2022 from <https://www.geeksforgeeks.org/bar-plot-in-matplotlib/>.

Honakan, Adiwijaya, and S. AL-Faraby. (2018). *Analisis Dan Implementasi Support Vector Machine Dengan String Kernel Dalam Melakukan Klasifikasi Berita*

Berbahasa Indonesia. vol. 5, no. 1, pp. 1701–1710.

KEMENKES (2021). *Unit Pelayanan Kesehatan Kementerian Kesehatan RI*. [online] Kemkes.go.id. Retrieved 29 Juni 2022 from

<https://upk.kemkes.go.id/new/4-manfaat-vaksin-covid-19-yang-wajib-diketahui>
Kirk R.E., and Othmer, D.F. (1996). “*Encyclopedia of Chemical Technology*”. vol.17, 4th edition, John Wiley & Sons Inc.,New York

L. Fadilla¹, A. Rizal², and E. Rachmawati³. (2011). Implementasi dan Analisis Content-Based Image Retrieval Pada Citra X-Ray Menggunakan Algoritma Hierarki dan Algoritma Fast Genetic K-Means. *Tugas Akhir- 2011 Fakultas Teknik Informatika Program Studi S1 Teknik Informatika*.

N. Wiliani, A. Sani, and A. T. Andyanto. (2019). Klasifikasi Kerusakan Dengan Jaringan Syaraf Backpropagation Pada Permukaan Solar Panel. *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*. vol. 5, no. 1, pp. 89–94, doi: 10.33480/jitk.v5i1.662.

P. N. Tan, M. S. (2006). *Introduction to Data Mining*. Boston: Pearson Education. Retrieved 2 Juli 2022 from <https://socs.binus.ac.id/2017/03/09/clustering/>.

Pradita A, Sulistianingsih D. (2021). *Urgensi Edukasi Program Vaksinasi Covid-19 Berdasarkan Peraturan Menteri Kesehatan Nomor 10 Tahun 2021*. *Jurnal Pengabdian Hukum Indonesia (Indonesian Journal of Legal Community Engagement) JPHI*. Retrieved 2 Juli 2022 from <https://journal.unnes.ac.id/sju/index.php/JPHI/article/view/49863/20129>

Riedel S, Morse S, Mietzner T, Miller S. Jawetz, Melnick, & Adelberg's. (1954, Aug). *Medical Microbiology*. 28th ed. New York: McGrawHill Education/Medical. p.617-22. Retrieved 2 Juli 2022 from <http://www.mcgrawhill.co.uk/html/0071476660.html>.

Sujarweni, V. W. (2014). *Metode Penelitian: Lengkap, Praktis, dan Mudah Dipahami*. Yogyakarta: Pustaka Baru Press.

Teguh Firmansyah (2021). *Khofifah: Cakupan Vaksin di Jatim Tertinggi*. Retrieved 2 Juli 2022 from <https://www.republika.co.id/berita/qwmwz4377/khofifah-cakupan-vaksin-di-jatim-tertinggi>.

Wu Z, McGoogan JM. (2020, Jan). Characteristics of and Important Lessons From the Coronavirus Disease 2019 (Covid-19) Outbreak in China. *Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention*. Available at : DOI: 10.1001/jama.2020.2648.

Zins, C. (2009). *What is the meaning of data information and knowledge*.