IMPLEMENTATION OF K-MEANS ALGORITHM FOR CLUSTERING OF COVID-19 VACCINATION IN EAST JAVA WITH ORANGE

Michael Sitorus[-1], Cornelia Antonieta DC.[-2*)], Cyntia Larasati[-3]

Information System
BRI Institute
http://bri-institute.ac.id/
michael.sitorus@bri-institute.ac.id[1], corneliaantonietadacosta15@gmail.com[2*)], clarasati099@gmail.com[3]
(*) Corresponding Author

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 penularan, mengurangi dampak berat dari virus, hingga mencapai kekebalan imunitas 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 didukung dengan gerakan memakai masker, menggunakan hand sainitizer, rajin mencuci tangan dan penelusuran kontak dari kasus positif. Penelitian ini bertujuan untuk mengklasterisasi secara demografis pelaksanaan vaksinasi di seluruh wilayah provinsi Jawa Timur pada tahun 2021. Metode yang digunakan dalam melakukan klastering ini adalah algoritma K-Means menggunakan tools yaitu Orange. 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.

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

Abstract

Entering the era of the Covid-19 pandemic, the government has intensively implemented a vaccination program until now. Covid-19 vaccination programs are implemented to strengthen the immune system, reduce the risk of infection, reduce the severe effects of the virus, and achieve herd immunity. In the implementation, the Covid-19 vaccination is regulated by the regional government in each province with a policy that requires vaccinating Covid-19 twice for everyone with specific criteria supported by movements to wear masks, use hand sanitizers, and diligently wash hands and contact tracing of positive cases. This study aims to demographically cluster the implementation of vaccinations in all areas of East Java province in 2021. The method used in conducting this clustering is the K-Means algorithm using tools, namely Orange. 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.

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

INTRODUCTION

Coronavirus is an RNA virus with a particle size of 120- 160 nm. Viruses these main infect animals, including among them are bats and camels. Before it happened to plague Covid-19, two types of coronavirus could infect humans: alphacoronavirus and betacoronavirus (Stefan Riedel, Stephen A. Morse, Timothy Mietzner, 2019). Since the first case in Wuhan, occur enhancement cases of Covid-19 in China every day and peak between the end of January and the beginning of February 2020. Originally most reports came from Hubei and surrounding provinces, then increased to other areas. Ain province and the whole of China (Wu, Z., & McGoogan, 2020).

Covid-19 infection can be transmitted through splashes of saliva (droplet) issued when the confirmed person Covid-19 experiences sneezing, coughing, or talking. Therefore, the splash of saliva can stick to objects so that if somebody touches a surface object inside it, there is a particle Covid-19

from the person confirmed so that it could cause transmission of Covid-19.

Factors that cause the high incidence of Covid-19 consist of internal factors (originating from within the individual), including knowledge, attitudes, and behavior, and external factors (coming from outside the individual), including government policies related to Covid-19 and the living environment (Mayasari, O. P., Ikalius, I., & Aurora, 2021).

1. Internal Factors; Internal factors are factors that exist among individuals who are facing the Covid-19 pandemic. Internal factors are divided into several indicators as follows:

a. Knowledge About Covid-19; Knowledge of cognition is essential in human behavior (Nur, 2021). Meanwhile, what is meant by knowledge of Covid-19 patients can be interpreted as the result of knowing from the patient about his illness, understanding his disease, ways of prevention, treatment, and complications (Mona, 2020).

b. Attitude in preventing Covid-19; Attitude is an evaluation or reaction to feelings. A person's attitude toward an object is supportive or prejudiced, as well as a feeling of being unfavorable or fair to the object. Attitudes toward preventing the transmission of Covid-19 are actions taken to reduce the risk of the Covid-19 disaster.

c. Behavior/Action; behavior is an activity of the human. The behavior referred to in this study is human activity related to preventing the Covid-19 disaster. A person's attitude toward an object is supportive or prejudiced, as well as a feeling of being unfavorable or fair to the object. (Hartanto, B., Sudrajat, D., & Badriatin, 2021). A person's behavior towards illness or disease is by the level of disease prevention, namely:

- Health promotion behavior
- Health preventive behavior
- Health seeking behavior
- Health rehabilitation behavior

2. External Factors; External factors are factors that exist outside the individual. External factors that influence the high incidence of Covid-19 are as follows:

a. Government Policies Related to Covid-19 prevention. The Indonesian government, led by Joko Widodo, has formulated several policies. There were at least nine legal products for the Covid-19 handling, namely four Presidential Decrees, two Presidential Regulations, one Government regulation, one Presidential Instruction, and one Government regulation instead of Law (Roziqin, A., Mas'udi, S. Y. F., & Sihidi, 2021).

b. Living environment; the living environment is one of the factors affecting public compliance with Covid-19 health protocols (Hamdani, 2020). Environmental factors that may underlie many Covid-19-19 cases include a community or individual disobedience in implementing health protocols, public indifference to the Covid-19 virus, and community activism.

The handling and prevention of this pandemic case have been carried out in various ways. For example, the implementation of 3M, namely using masks, washing hands with soap, and maintaining a minimum distance of 1 meter, is critical as a form of essential protection to prevent the spread of Covid-19 (Eriyani, T., Shalahuddin, I., & Rosidin, 2021). Not only that, but giving vaccine also aims to bring up the response immunity body somebody to attack virus SARS-Covid-19. Vaccination is the process of providing immunizations by hand injection, or it can be dripped into the mouth, which will later function to increase the production of antibodies in the body that are useful for warding off certain diseases (Larasati, P. A., & Sulistianingsih, 2021) so that the body could oppose the infection virus Covid-19. Naturally, system immunity body to Covid-19 after being vaccinated does not participate immediately could be formed by instant, protocol health 3M which planned government must permanent held for provide maximum protection against Covid-19 attacks (Kemenkes RI, 2021).

One of the implementations of the Covid-19 vaccination is in East Java Province. East Java Governor Khofifah Indar Parawansa said East Java's Covid-19 program vaccination coverage is the highest compared to other provinces in Indonesia. Khofifah is urging residents who have not yet been vaccinated to join the vaccination program. According to Task Force data Handling Covid-19, the number of residents who have got injection dose first vaccine Covid-19 in East Java is as many as 7,058,237 people, more than in DKI Jakarta (6,693,688 people), West Java (5,134,735 people), Java Central (4,523,284 people), and Bali (2,946,919 people). The inhabitant who has two vaccine injections or has finished undergoing vaccination in East Java recorded as much 2,694,731 people, more than in West Java (2,315,419 people), Central Java (2,311,019 people), DKI Jakarta (2.041.918 person), and Bali (783,613 people). East Java Governor optimistic with scope vaccination which has already achieved East Java can realize communal immunity against Covid-19 in August 2021 (Firmansyah, 2021). In case this with the availability of raw data on Covid-19 vaccination in East Java province that

can be processed, this research purposed increase knowledge for Covid-19 vaccination, help in seeing the spread and development of vaccination Covid-19 in East Java Province use clustering method K-Means with Orange tools, and provide recommendations to the government regarding the optimization of vaccination, especially in the province of East Java.

RESEARCH METHODS

This study used a quantitative research method. Quantitative methods systematically identify the value of key multi-stakeholders and their multiple interactions, providing a unified strategy for promoting value creation among stakeholders (Zheng, X., Lu, Y., Li, Y., Le, Y., & Xiao, 2019). Implementation study this there is several stages are carried out, namely starting from raw input data, K-Means which consist of bar plots and predictions (clustering), as well as distributions like in Figure 1 :
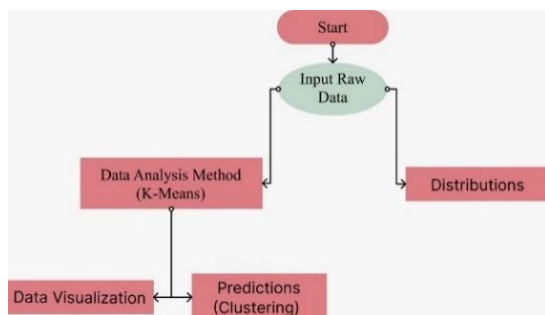

Figure 1. Workflow

The raw data study originated from Portal Service Health Government East Java which serves data vaccination Covid-19 area East Java https://infocovid19.jatimprov.go.id//vaksinasi on January – March 2021.

As for the process conducted in a study with a workflow like in Figure 2 following:
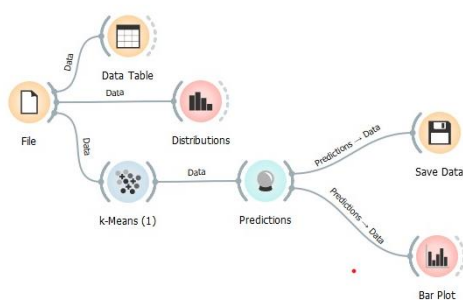

Figure 2. Workflow Clustering Vaccination Covid-19 in East Java

Input Raw Data

Raw data is collected from data vaccination Covid- 19 regions of East Java in January – March. Where inside data, there is: (1) City or Regency, (2) Vaccinate, (3) Fully Vaccinated, and (4) Total.

Data Analysis Method

The data analysis method uses K-Means, an algorithm to group multiple data. K-Means is a non-hierarchical data grouping technique that allows dividing data into two or more groups (Gustientiedina, G., Adiya, M. H., & Desnelita, 2019). This algorithm uses an iterative process to get the database clusters. Besides that, algorithm K-Means need amount cluster beginning which wanted to become input and produce centroid point end as outputs. The following is the formulation of K-Means:

$$D_e = \sqrt{(x_{i-}s_i)^2 + (y_{i-}t_i)^2} \dots\dots\dots\dots\dots\dots\dots (1)$$

Where :
- D_e is Euclidean Distance
- i is an object
- (x,y) are the coordinate object
- (s,t) centroid coordinates.

Data Visualization

Data visualization can be defined in several ways. Most definitions focus on the connection between data and computer technology for converting data into visual or sonic form. Process basic data visualization in any form can be changed into a graphic image (Madyatmadja, E. D., Nuramalia, A. N., Kusumawati, L., Jamil, S. P., & Kusumawardhana, 2021). The author uses a bar plot type diagram to visualize the data in this study.

Clustering

Clustering is the process of partitioning a set of data objects into a set part called clusters, where the object has similar characteristics between the same as each other and are different from the cluster. A cluster is modeled as a statistical distribution, i.e., objects are generated by a random process from a statistical distribution characterized by several statistical parameters, such as the mean and variance. This viewpoint generalizes the notion of a prototype and enables well-established statistical techniques (Pang-Ning Tan, Michael Steinbach, Vipin Kumar, 2018). Process this clustering is used to help predict and analyze the success of the Covid-19 vaccination in the East Java area from January – March 2021.

Distributions

Data were obtained from the results study and processed by needs. The frequency distribution is a series of numerical data according to quantity or quality. A data series of numbers according to their quantity is called a frequency distribution quantitative data; on the other hand, data arranged according to quality is called a frequency distribution qualitative (Wahab, A., Syahid, A., & Junaedi, 2021).

RESULTS AND DISCUSSION

With the used methodology which has displayed in the previous chapter, the results of the analysis and calculations were obtained as follows:
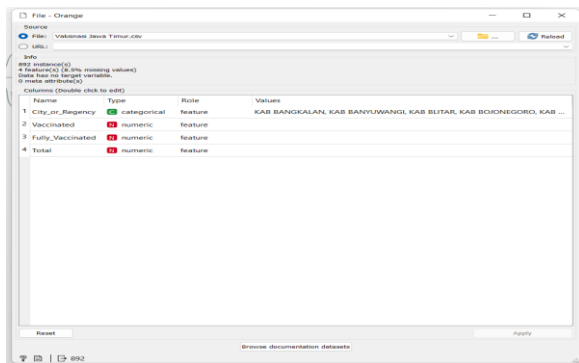
Input Raw Data

Figure 3. *Input Raw Data* on Orange

Based on Figure 3, it can be seen that the raw data on Covid-19 vaccination in East Java has four attributes that are:
1. City_or_Regency contains East Java cities or counties with a categorical data type.
2. Vaccinated including public set already accepting vaccination for Covid-19 in East Java with the numeric data type.
3. Fully_Vaccinated contains the number of people vaccinated against Covid-19 in East Java, and the type data is numeric.
4. Total population fully or partially vaccinated against Covid-19 in East Java with numerical data.

Implementation K-Means
1. Centroid Data

In the application algorithm, K-Means generated the value of the midpoint or centroid of the data obtained. Cluster determination is divided into three parts based on the calculation by Orange tools like in Table 1, namely low cluster (C1), middle cluster (C2), and high cluster (C3). Then the value of the middle value there has three points. Cluster point determination is done with the take score smallest (minimum) for the low cluster (C1), the average value (average) for a medium cluster (C2), and the largest value (maximum) for a high cluster (C3). The point value can be counted by formulation following :

$$\overrightarrow{V}_{ij} = \frac{1}{N}\sum_{k=0}^{N} x_{kj} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots(2)$$

Where :
- $\overrightarrow{V}_{ij}$ is the centroid/average of the j-th cluster for the j-th variable
- N is the amount of data that is a member of the i-th cluster
- i,k is the index of the cluster j is the index of the variable
- x_{kj} is the k-th data value in the cluster for the j-th variable

And the result is in the table following :

Table 1. Centroid Data (Iteration 1)

	Vaccinated	Centroid
Min (C1)	4821	0.489175
Average (C2)	3068	0,556051
Max (C3)	34316	0,685673

Based on Table 1, the centroid value has a value range of 1 to -1, if the value listed is closer to 1, the better the cluster results.

2. Visualization Data

On this data, visualization conducted is used chart bar plot with score values vaccinated and group by the cluster as well as the annotation none with visualization as follows.
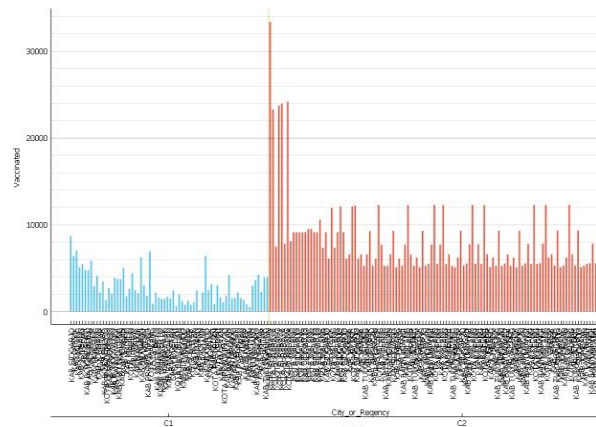

Figure 4. Visualization data on Orange

3. Clustering Data

Centroid allows clustering data into three clusters. A processing cluster with the smallest possible distance to all processed data. From data amount county town vaccinated Covid-19 on East Java Province obtained grouping on the first iteration for three clusters. Following Table 2, the cluster with the lowest value in the implementation of Covid-19 vaccination in East Java (C1) is Pasuruan Regency. The cluster with a middle value (C2), namely Kediri City, and the cluster with the highest value (C3) is Surabaya City.

Table 2. Grouping Data Clustering

Cluster	Silhouette	City_or_Regency	Vaccinated	Fully Vaccinated	Total
C1	0,489175	Kabupaten Pasuruan	4821	4282	9103
C2	0,556051	Kota Kediri	3068	0	3068
C3	0,685673	Kota Surabaya	34316	27489	61805

Process K-Means will continue until the grouping data is the same as the previous iteration. In the first iteration, we obtained cluster data amount vaccination Covid-19 based on county town. The iteration process stops at iteration; on the next iteration, a method is conducted to look for the middle value. After getting the middle value, the same process is done by finding the distance closest. Process search distance shortest.

5. Distribution Data

Based on Figure 5, data grouping done to three clusters with one iteration obtained the same result, and distribution could view on the chart following:
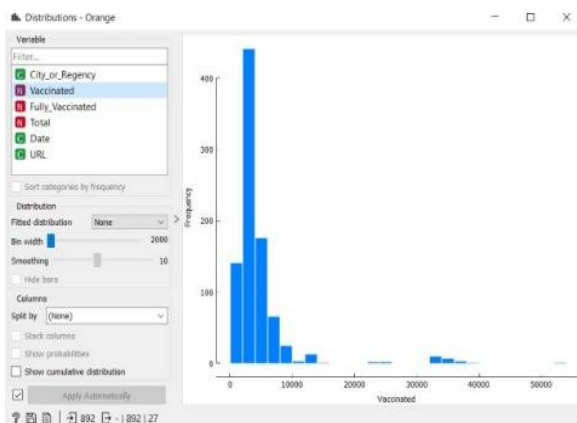

Figure 5. Distribution data on Orange

6. Results Implementation by Orange

In clustering, the algorithm K-Means is used for group data based on attributes. Iterations mean the number of times the algorithm is run from random starting positions. The result with the smallest within-cluster sum of squares is used; max iterations is the maximum number of iterations within each algorithm run that can be set manually. Application orange will help process clustering and is shown in the following Figure 6 :
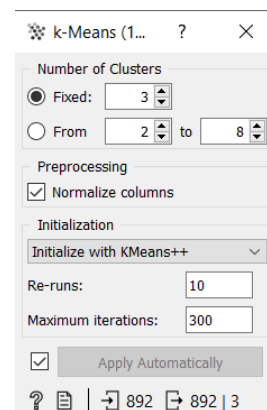

Figure 6. Results implementation tool orange

From the picture in Figure 6, information is obtained that there are three clusters, low cluster (C1), medium cluster (C2), and high cluster (C3). By normalizing the columns and initializing K-Means.

CONCLUSIONS AND SUGGESTIONS

Conclusion

By using the orange tool, we get the results that the regions in East Java Province that have successfully implemented the Covid-19 vaccination with the highest success are the city of Surabaya. The City of Kediri is an area with a moderate success rate, and Pasuruan Regency is included in a room with low success.

Suggestion

Findings from this study may provide information and suggestions to the government, especially Pasuruan Regency, which has the lowest implementation of Covid-19 vaccination in East Java, to resume further actions soon from this study may provide information and suggestions to the government, especially Pasuruan Regency, which has the lowest implementation of Covid-19 vaccination in East Java, to resume further actions soon. For example, through socialization, education,

vaccination rewards, and more significant out-of-town travel requirements for those who have not been fully vaccinated.

REFERENCES

Eriyani, T., Shalahuddin, I., & Rosidin, U. (2021). Edukasi 3M Dalam Meningkatkan Self-Awareness Terhadap Penyebaran Covid-19 Di Smkn 4 Garut. *Kumawula: Jurnal Pengabdian Kepada Masyarakat*, *4*(1), 98. https://doi.org/10.24198/kumawula.v4i1.32341

Firmansyah, T. (2021). *Khofifah: Cakupan Vaksin di Jatim Tertinggi*. Republika. https://www.republika.co.id/berita/qwmwz4377/khofifah-cakupan-vaksin-di-jatim-tertinggi.

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

Hamdani. (2020). *Kepatuhan Sosial di Era New Normal*. https://www.ajnn.net/news/kepatuhan-sosial-di-era- new-normal/indeks.html

Hartanto, B., Sudrajat, D., & Badriatin, T. (2021). Community Behavior in Health Protocol Policy Perspective Post Covid-19 Vaccination. *Sawala: Jurnal Administrasi Negara*, *9*(2), 192–200. https://doi.org/10.30656/sawala.v9i2.3966

Kemenkes RI. (2021). *4 Manfaat Vaksin Covid-19 yang Wajib Diketahui*. Kementrian Kesehatan RI. https://upk.kemkes.go.id/new/4-manfaat-vaksin-covid-19-yang-wajib-diketahui#:~:text=Mengurangi Risiko Penularan,virus dan mengurang risiko terpapar.

Larasati, P. A., & Sulistianingsih, D. (2021). Urgensi Edukasi Program Vaksinasi Covid-19 Berdasarkan Peraturan Menteri Kesehatan Nomor 10 Tahun. *Jurnal Pengabdian Hukum Indonesia*, *4*(1), 99–111. https://journal.unnes.ac.id/sju/index.php/JPHI/article/view/49863/20129

Madyatmadja, E. D., Nuramalia, A. N., Kusumawati, L., Jamil, S. P., & Kusumawardhana, W. (2021). Data Visualization of Internet Usage in The Jabodetabek Area. *Infotech: Journal of Technology Information*, *7*(1), 55–62.

https://doi.org/10.37365/jti.v7i1.108

Mayasari, O. P., Ikalius, I., & Aurora, W. I. D. (2021). Faktor-Faktor yang Berhubungan dengan Perilaku Masyarakat dalam Pencegahan Covid-19 di Wilayah Kerja Puskesmas Kenali Besar Kecamatan Alam Barajo. *Medical Dedication (Medic): Jurnal Pengabdian Kepada Masyarakat FKIK UNJA*, *4*(1), 146–153. https://doi.org/10.22437/medicaldedication.v4i1.13467

Mona, N. (2020). Konsep Isolasi dalam Jaringan Sosial untuk Meminimalisasi Efek Contagious (Kasus Penyebaran Virus Corona di Indonesia). *Jurnal Sosial Humaniora Terapan*, *2*(2), 1–10. https://doi.org/10.7454/jsht.v2i2.86

Nur, S. A. & E. S. (2021). Hubungan Tingkat Pengetahuan dengan Perilaku Kesehatan Reproduksi Remaja pada Siswa SMK Kabupaten Semarang. *Jurnal Ilmiah Kesehatan Ar-Rum Salatiga*, *5*(2), 45–52.

Pang-Ning Tan, Michael Steinbach, Vipin Kumar, A. K. (2018). *Introduction to Data Mining* (2nd ed., Issue December). Pearson.

Roziqin, A., Mas'udi, S. Y. F., & Sihidi, I. T. (2021). An analysis of Indonesian government policies against COVID-19. *Public Administration and Policy*, *24*(1), 92–107. https://doi.org/10.1108/PAP-08-2020-0039

Stefan Riedel, Stephen A. Morse, Timothy Mietzner, S. M. (2019). *Medical Microbiology* (28th ed.). Mc Graw Hill.

Wahab, A., Syahid, A., & Junaedi, J. (2021). Penyajian Data Dalam Tabel Distribusi Frekuensi Dan Aplikasinya Pada Ilmu Pendidikan. *Education and Learning Journal*, *2*(1), 40–48. https://doi.org/10.33096/eljour.v2i1.91

Wu, Z., & McGoogan, J. M. (2020). 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. *JAMA (Journal of the American Medical Association)*, *323*(13), 1239. https://doi.org/10.1001/jama.2020.2648

Zheng, X., Lu, Y., Li, Y., Le, Y., & Xiao, J. (2019). Quantifying and visualizing value exchanges in building information modeling (BIM) projects. *Automation in Construction*, *99*(November 2018), 91–108. https://doi.org/10.1016/j.autcon.2018.12.001