International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Regional Clustering Model of Covid-19 Cases at the Early of the Pandemic in Indonesia

Arief Wibowo*[1], Andy Rio Handoko[2], Muhammad Zarlis[3]

Submitted: 23/08/2023 **Revised:** 11/10/2023 **Accepted:** 24/10/2023

Abstract: President of the Republic of Indonesia Ir. Joko Widodo announced the first case of the COVID-19 virus in Indonesia on March 2, 2020. 2019-nCoV, the original name of this coronavirus, is becoming increasingly well known because it has become a pandemic in more than 60 countries in just two months since it was discovered. Only one month since this virus was found in Indonesia, almost 1,800 people have tested positive for Covid-19. DKI Jakarta Province is the epicenter of this virus pandemic, with the number of positive sufferers of more than 5600 people as of May 14, 2020. With this vast spread, DKI Jakarta was designated as the first province to implement Large-Scale Social Restrictions (PSBB) based on the Decree of the Minister of Health of the Republic of Indonesia as of April 7, 2020. This research aims to model the clustering of sub-district areas with positive COVID-19 cases in DKI Jakarta. The model was built using the K-Means algorithm, a clustering data mining algorithm that efficiently forms clusters from the past data learning process. The research data comes from the official Corona unique website in Jakarta, namely http://corona.jakarta.go.id, which was cross-tested with other sources, accompanied by elaboration with regional demographic data from the DKI Jakarta Provincial Government website. The research results show that clustering or grouping of areas in DKI Jakarta with the number of positive sufferers is formed at the value k=2 with a Davies Bouldin Index evaluation value of 0.182. The clustering pattern visualized as an electronic-based map can be connected to data on the favorable distribution of COVID-19 in sub-district areas in DKI Jakarta. Jakarta is the epicenter of the pandemic in Indonesia, so this research is important to find out how the country's capital region is clustered to face other pandemics in the future. The research results can be used by stakeholders in making decisions related to handling the COVID-19 pandemic, especially in the DKI Jakarta area, based on the regional clusters formed.

Keywords: *Clustering, Data Mining, Mapping COVID-19 Cases*

1. Introduction

The spread of the COVID-19 virus in Indonesia occurred quickly. In just one month since it was first announced on March 2, 2020, the number of COVID-19 sufferers in Indonesia has almost reached two thousand people across thirty-two provinces. As of April 2, 2020, at 17.00 WIB, 7,425 people were examined with examination results, namely 5,635 negative people and 1,790 positive confirmed cases of COVID-19, with a record that 112 patients recovered and 170 people who died [1].

Based on April 5, 2020 data from the Ministry of Health of the Republic of Indonesia, the distribution areas in Indonesia are evenly distributed in almost all provinces from Aceh to Papua with 2,273 people. From these data, on the island of Java, it turns out that DKI Jakarta Province has the most significant cumulative number of patients with more than one hundred people, followed by West Java Province with 252 people, Banten with 177 people, East Java with 188 people, Central Java with 120 people and DI Yogyakarta with 34 people. Official information recorded that there were at least 1,124 people who tested positive for COVID-19 in the capital city of Jakarta as of April 5, 2020 [2].

DKI Jakarta, the capital city of Indonesia, has become the epicenter of the spread of the coronavirus, which does not yet have an antidote. The distribution area is almost evenly distributed throughout Jakarta; at least there are positive patients in five administrative city areas of DKI Jakarta (Jakarta). With the spread so fast, handling the COVID-19 pandemic has triggered the Provincial Government of DKI Jakarta to propose establishing Large-Scale Social Restrictions (PSBB). On April 7, 2020, the Minister of

1 Faculty of Information Technology Universitas Budi Luhur, INDONESIA

ORCID ID: 0000-0001-8544-5145

2 Faculty of Information Technology Universitas Budi Luhur, INDONESIA

ORCID ID: 0009-0009-5081-8760

3Information System Management Department, BINUS Graduate Program – Master of Information System Management, Bina Nusantara University, INDONESIA.
ORCID ID: 0000-0003-0520-7273

** Corresponding Author Email: arief.wibowo@budiluhur.ac.id*

1arief.wibowo@budiluhur.ac.id,2andy.handoko@budiluhur.ac.id, 3muhammad.zarlis@binus.ac.id

Health, by his authority as regulated in Law no. 6 of 2018 concerning Health Quarantine, granted the Governor of DKI Jakarta's request for the proposed establishment of PSBB in the Jakarta area. On the date of this determination, at least in the Jakarta area, the number of positive COVID-19 patients was recorded as many as 1,395 people, consisting of 827 people who had been identified based on their urban village domicile and who had not been identified.

This research is an initial study that conducts clustering modeling of urban villages in Jakarta that have some positive cases of COVID-19 since the first case was announced until the implementation period of the Large-Scale Social Enforcement (PSBB) policy by the Provincial Government of DKI Jakarta. The model is built with a data mining approach, producing a clustering pattern based on past data learning processes. The expected results of the research are the discovery of a way of regional grouping according to the number of positive cases of COVID-19 and based on population as well as regional density demographic data per District *(Kelurahan)* in Jakarta. The clustering pattern will be visualized as an electronic-based map connected to data on the distribution of positive cases of COVID-19 in Jakarta Province.

Previous studies that have been carried out include clustering modeling for areas with the spread of Dengue Fever, resulting in three clusters. Primary data of DHF sufferers per village, without collaboration with other data, and the pattern of modeling results is only visualized textually [3]. In forming product clusters sold at supermarkets, with two groups and no other data collaboration, the pattern of modeling results is only presented textually [4].

Another study conducted modeling of KIP holders to form clusters in the categories of rich, simple, and low-income families. The research contribution lies in the testing process using 200 factual data, with an accuracy of 69% [5]. Meanwhile, area-based clustering modeling was also carried out to determine the spread of Diarrheal disease. This study uses primary data of diarrhea sufferers per sub-district, without collaboration with other data, and the pattern of modeling results is only visualized textually. In this study, data transformation was carried out for the number of sub-districts in one sub-district to become a specific value scale.

Studies on other clustering modeling have also been carried out with one algorithm for MSME data based on the latest assets and turnover in an area and their suitability with the potential of the existing site. Information is presented in the form of text and geographic-based maps. The data source is a collaboration of various origins. However, the presentation of the results has yet to be integrated between the Google-based clustering maps built with the MSME area potential maps that support them [6]. At the same time, the grouping of sub-districts is based on possible agricultural products to optimize the implementation of government programs in agriculture. The clustering results are presented only in the form of narrative or textual without integrating other data/information.

This study will form a regional clustering pattern using data on the distribution of positive COVID-19 cases in DKI Jakarta Province. The difference in this study is the involvement of demographic data in population density and area to support the specification of the results of the clusters formed. In addition, this study will visualize the clustering results in an online map, namely Google Maps, which presents information dynamically from the effects of clustering modeling, which is strengthened by secondary data. This proposed contribution aims to enrich the outputs and research objectives so that the research can be completed relatively quickly. This research is expected to be helpful for the decision-making process during the COVID-19 pandemic in DKI Jakarta and in Indonesian regions in general.

One of the challenges and novelties in this research is using primary data during the critical period of the spread of COVID-19 in Jakarta, especially one month after the case was first discovered in Indonesia. These initial conditions illustrate the preparedness situation of the Jakarta region for the initial conditions of the pandemic in Indonesia.

2. Research Methodology

In this research, the method used is clustering modeling using the K-Means algorithm. Modeling experiments were carried out to obtain clusters with the best evaluation index, referring to previous studies on clustering data mining [7]. This research stage consists of five steps: data collection, initial data processing, proposed methods, model experiments, testing, and evaluation and validation of results. we have chosen the k-means algorithm because of its suitability to the type of data available, based on previous research [8], [9]. K-means has high advantages, especially in processing numerical and text data, this is reinforced by previous studies which analyzed k-means comprehensively [10]. Even K-means can be optimized to achieve more optimal performance [11].

Secondary data in the research is a combination of data on the distribution of COVID-19 cases in DKI Jakarta from the page https://corona.jakarta.go.id/id, which will be cross-checked with the data on the page https://jeo.kompas.com. Research data also comes from the Jakarta Provincial Government website on the page http://data.jakarta.go.id, in the form of population data and regional density demographics.

This research methodology refers to the CRISP-DM model, which stands for Cross Industry Standard Process for Data Mining. CRISP-DM is a data mining standardization prepared by three initiators of the data mining market, according to Larose, as quoted from a previous study [12]. In this research, the cycle for the data mining process was divided into six stages, where the dependencies between each step are depicted with arrows. Figure 2 shows the research methodology referred to based on the CRISP-DM method."

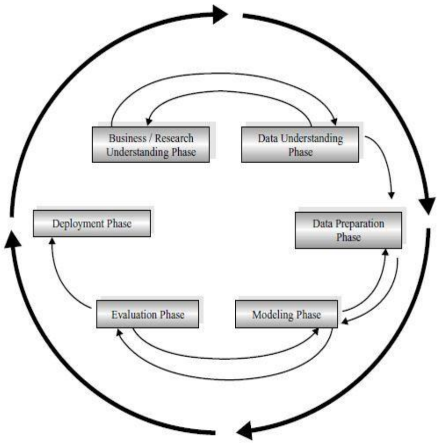

Fig. 1. CRISP-DM Methodology.

Based on Figure 1, the outermost cycle circle illustrates that data mining with the CRISP-DM methodology can take experience from past projects to provide input for new projects. Some literature describes that from the evaluation stage, you can send the analysis back to one of the previous stages. However, for simplicity, this literature describes the most common process where, from the evaluation stage, you can return to the modeling stage.

At the Business Understanding stage, the focus is on understanding the project's goals. From a business perspective, it needs to turn this into a data mining problem and an initial plan to achieve these goals. Activities carried out include:

- determining overall objectives and requirements
- translating these goals
- determining limitations in formulating data mining problems
- preparing an initial strategy to achieve these goals

At the Data Understanding stage, data is collected and then studied to get to know the data, identify and know the quality of the data, and detect exciting subsets of the data that can be used as hypotheses for hidden information. In the Data Preparation Stage, preparation is carried out regarding the data that will be used in the next stage. Activities carried out include:

- selecting cases and parameters to be analyzed
- carrying out transformations on specific parameters, and
- cleaning the data to be ready for the modeling stage

The Modeling stage determines the appropriate data mining techniques, tools, and algorithms to apply. Suppose it is necessary to adjust the data to specific data mining techniques. In that case, you can return to the data preparation stage. At the Evaluation stage, the data mining results produced in the modeling process in the previous step are interpreted. Evaluation is carried out on the model applied in the last scene with the aim that the model is determined by the objectives to be achieved in the first stage. The final step is Deployment, namely preparing a report on the results obtained from the evaluation in the previous setting or from the data mining process carried out as a whole.

Referring to Knowledge Discovery in Databases, all stages of this research were carried out while remaining within the CRISP-DM framework. The initial work of this study was data pre-processing. The data is processed into several groups or clusters using the K-Means algorithm. After the cluster rules have been formed, the information is calculated using the Euclidean distance formula to calculate the distance between data so that group members can be determined from each cluster. However, all modeling experiments were completed using the developed clustering modeling prototype application, and testing of the cluster model formed was evaluated using the Davies Bouldin Index (DBI) parameters.

3. Results and Discussion

This section will explain the stages of problem-solving with the CRISP-DM approach. The problem-solving steps carried out start from the research understanding phase, data understanding phase, data preparation phase, modeling phase, evaluation phase, and development phase.

Data from the research understanding phase ensures that the researcher prepares measurable, systematic, and scientific problem-solving steps based on theoretical or literature studies. It is confirmed that the research stage will start with data acquisition to be understood and analyzed first before processing. Researchers conducted a relatively limited literature review online from various data sources to understand the business processes in the COVID-19 pandemic case, which has been going on for about six months since the issue was discovered. For the data acquisition process to be accurate, researchers have ensured the availability of valid and reputable primary data sources, from the site http://corona.jakarta.go.id and other trusted secondary data sources, namely respected mass media that present data and information about the spread of positive cases of COVID-19 in DKI Jakarta as a comparison of the validity of the data acquired. Researchers believe both data sources have content that can be used for the data understanding phase.

When this research was conducted, another national data source that could be accessed was the Ministry of Health's website. Central data sources such as the Central Statistics Agency still need to provide the required data; therefore, data acquisition is only carried out online due to time constraints and safety reasons due to social restrictions. In the end, the researcher believes that the business process in the current preliminary study starts from data observation, data acquisition, data analysis, and data understanding, to be then processed with other research stages after the data is understood.

After the researcher understands the type of data acquired, the next phase is to understand the data that has been received. Data acquisition was carried out repeatedly, according to the conditions of the research object, namely the spread of positive cases of COVID-19 in DKI Jakarta. To understand the sequence or chronology of events in the space of COVID-19 along with the data that appears, a timeline or timeline of events during the spread of positive cases of COVID-19 in DKI Jakarta was carried out. By mastering this timeline-based data, researchers can select or select data or groups of data as input in the modeling experiment process being carried out. The data collection timeline is shown in Figure 2.

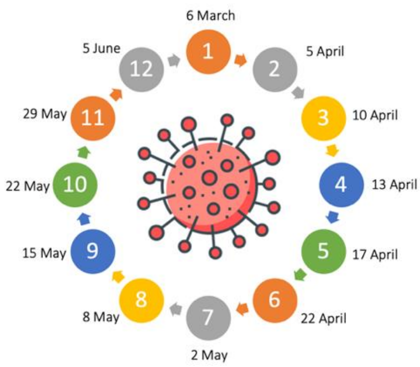

Fig. 2. Data Collection Timeline

Research data was collected from data on the distribution or number of positive cases in DKI Jakarta Districts *(Kelurahan)* during the specified period, namely from March 6 to June 5, 2020. This data was supplemented with secondary data from the DKI Jakarta Provincial Government's Integrated Data Portal in the form of area data per sub-district in DKI Jakarta and population density per square KM. This research requires periodic data collection to adjust developments in the distribution of positive cases in DKI Jakarta and policies issued by the Provincial Government.

It can be seen in Figure 4 that the secondary data collection timeline consists of twelve stages. The first stage, on March 6 2020, was data collection for the names of sub-districts on the website http://corona.jakarta.go.id as the main reference page for collecting data for this research.

This website was inaugurated on March 6, 2020, by the DKI Provincial Government to provide initial information about handling and anticipating the presence of the COVID-19 virus in Indonesia, especially for residents of DKI Jakarta [13]. The launch of this website is a follow-up to the first COVID-19 case in Indonesia, which was announced on March 2, 2020, by the President of the Republic of Indonesia Ir. Joko Widodo [8], as Case 01. As of March 6, 2020, four positive cases of Covid were recorded in DKI Jakarta [14].

The second phase, April 5, 2020, was to collect further data on the spread of positive COVID-19 cases in DKI Jakarta. This phase is the momentum of the passage of one month since the space of positive instances of COVID-19 was first announced on March 2, 2020. The number of positive cases of COVID-19 in Jakarta on that date was 657 people. This data appears on the page http://corona.jakarta.go.id, which was observed on April 5, 2020, at 08:00 WIB.

The third phase, April 10, 2020, was the collection of further data on the spread of COVID-19 in DKI Jakarta, with the finding that the number of positive cases reached 1719 people nationally. On this date, the DKI Jakarta Government determined the implementation of the Large-Scale Social Restrictions known as PSBB policy from April 10 to April 23, 2020, based on Governor Regulation Number 33 of 2020 [15]. The number of positive cases of COVID-19 in DKI Jakarta on that date was 944 people. This data appears on the web page based on observations at 08:00 WIB.

The fourth phase, April 13, 2020, collects further data on the spread of positive COVID-19 cases in DKI Jakarta. It is known that the number of positive cases of COVID-19 in DKI Jakarta on this date was 1132 people.

The fifth phase, April 17, 2020, collects further data on the spread of positive COVID-19 cases in DKI Jakarta. This date is recorded as one week when the PSBB was implemented in DKI Jakarta. The number of positive cases of COVID-19 in Jakarta on that date reached 1609 people.

The sixth phase, April 22, 2020, collects further data on the spread of positive COVID-19 cases in DKI Jakarta. It is known that the number of positive cases of COVID-19 in DKI Jakarta on that date was 2128 people. This date is the last day of implementation of the DKI Jakarta Provincial Government's first PSBB policy. Seeing the spread of COVID-19 sufferers in Jakarta, which continued to increase, the DKI Provincial Government implemented the PSBB Phase 2 policy for 28 days, effective from April 24 to May 22, 2020 [16].

The seventh phase, May 2, 2020, collects further data on the spread of positive COVID-19 cases in DKI Jakarta. It is known that the number of positive cases of COVID-19 in DKI Jakarta on that date reached 3132 people. This is the 60th day since the first positive case of COVID-19 was discovered in Jakarta, the first case in Indonesia.

The eighth phase, May 8, 2020, collects further data on the spread of positive COVID-19 cases in DKI Jakarta. It is known that the number of positive cases of COVID-19 in DKI Jakarta on that date reached 3548 people.

The ninth phase, May 15, 2020, collects further data on the spread of positive COVID-19 cases in DKI Jakarta. It is known that the number of positive cases of COVID-19 in DKI Jakarta on that date reached 4114 people.

The tenth phase, May 22, 2020, was the collection of further data on the spread of positive COVID-19 cases in DKI Jakarta, using the observation method of visits to web pages carried out at 08:00 WIB. It is known that the number of positive cases of COVID-19 in DKI Jakarta on that date reached 4670 people. On this date, PSBB Phase 2 ended and continued with the PSBB Phase 3 policy, which took effect from May 23 to June 4, 2020 [17].

The eleventh phase, May 29, 2020, collects data on the spread of positive COVID-19 cases in DKI Jakarta. It is known that the number of positive cases of COVID-19 in DKI Jakarta on that date reached 5032 people.

The twelfth phase, June 5, 2020, collects final data on the spread of positive COVID-19 cases in DKI Jakarta. It is known that the number of positive cases of COVID-19 in DKI Jakarta on this date reached 5535 people. Today, the PSBB policy was declared to have ended. The DKI Provincial Government, through Governor's Decree (Kepgub) Number 563 of 2020, implemented the Transitional PSBB Phase, namely the Implementation of Stages and Implementation of Large-Scale Social Restrictions/Activities during the Transition Period towards a Healthy, Safe and Productive Society.

The impact of the implementation of the PSBB policy, which ended, was a change in the policy implemented in DKI Jakarta in the form of a Transitional PSBB period, namely allowing several fields and activities to be carried out by the community and starting to relax the ßPSBB rules. The state of relaxation includes the opening of several business activities in the office, trade, sports, and transportation sectors with tolerance in the form of activities that can be carried out at a maximum of 50% of the capacity of people in the activity/location/building or vehicle used [18]. Based on this timeline, June 5, 2020, is this study's final data collection stage, namely, when the PSBB policy in DKI Jakarta was completed.

The reason for stopping data collection during this phase is based on an analysis of the Transitional PSBB policy, which relaxes population activities and is thought to have caused the distribution of positive COVID-19 cases to tend to change compared to when the PSBB situation was implemented where activities in various sectors were stopped entirely. Thus, collecting research data during the PSBB period was deemed appropriate as secondary data for an initial study of grouping areas with positive COVID-19 cases in DKI Jakarta.

This research also used secondary data sources from the Integrated Data Portal of the DKI Jakarta Provincial Government in the form of Area Area and Density by Region. This data was obtained via the page http://data.jakarta.go.id [19]. The data understanding phase results in knowing the demographic data in the form of a profile resulting from data acquisition regarding the spread of positive COVID-19 cases in DKI Jakarta.

In the pre-processing stage, before the data is ready to be analyzed for problem-solving. The handling process is carried out on inconsistent, conflicting, and duplicate data. Inconsistent data is incomplete data being found and requires cross-checking to other data source pages, such as https://jeo.kompas.com. Cross-checking is carried out in stages according to the discovery of inconsistencies in the primary source data belonging to the Provincial Government. DKI Jakarta. The data pre-processing stage is also carried out to obtain summaries based on the conditions of administrative city areas in DKI Jakarta for presenting profile or demographic data.

The selection process was conducted on 267 lines of the research dataset with primary and secondary attribute dimensions. Primary attributes consist of data on the names of sub-districts and sub-districts and the number of positive cases occurring in each sub-district. In contrast, secondary characteristics consist of latitude and longitude coordinate data, regional density data per sub-district, and data on the area of sub-districts in Jakarta.

3.1. Clustering Modeling Phase

The Modeling section explains the steps taken, from implementing the k-means algorithm to model analysis and testing.

At this stage, an algorithm is selected for solving research problems that want to produce cluster models from graduate data based on the results of tracking studies. Based on the literature study used as a reference, the K-Means algorithm is the most suitable method for the current research data. The modeling in this study was completed using Rapidminer Studio software running on the Mac operating system (macOS).

At the modeling stage, the data is preprocessed and imported into the process design in Rapidminer, as shown in Figure 3.

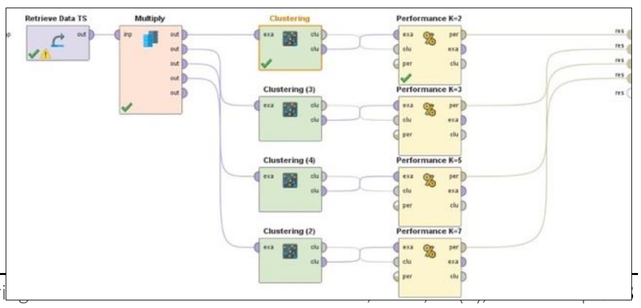

Fig 3. Clustering Modeling Design in Rapidminer Studio

It can be seen in Figure 1 that modeling is carried out simultaneously for various scenarios (multiply), namely clustering with K=2, K=3, K=4 and K=5 on the training dataset. In testing using the Davies Bouldin Index value, the best results were obtained from cluster k = 2. However, this study used two tests using two different datasets. A comparison of DBI values can be seen in Table 1.

Table 1. Comparison of Davies Bouldin Index Values

k	DBI	
	Dataset A	Dataset B
2	0.182	0.325
3	0.708	0.615
4	0.919	1.770
5	0.991	0.921

Once the number of selected datasets has been fulfilled, the modeling process can be done directly by applying the K-Means algorithm. The first work carried out is to select centroids randomly with an illustration of a portion of the data.

Based on the results of the modeling carried out, the final results of the clustering process were obtained using the k-means method combined with secondary research data in the form of area and population density data to get more specific cluster specifications. The final clustering results are shown in Table 2.

Table 2. Regional clusters formed

Cluster	Number of Cluster Members	Average Area	Average Area Density
0	4	1.71 km^2	9,100 citizen/km^2
1	6	1.42 km^2	36,653 citizen/km^2

It can be seen in Table 2 that Cluster 0 has a total of 3 cluster members and an average area of 1.71 km^2 and an average density of 9,100 citizen/km^2. Cluster 1 has a total of 7 members and an average area of 1.42 km^2 and an average density of 36,653 citizen/km2.

Based on the model formed, the development phase is the last phase carried out in this study. The results of developing the application prototype can be seen in Figure 4.

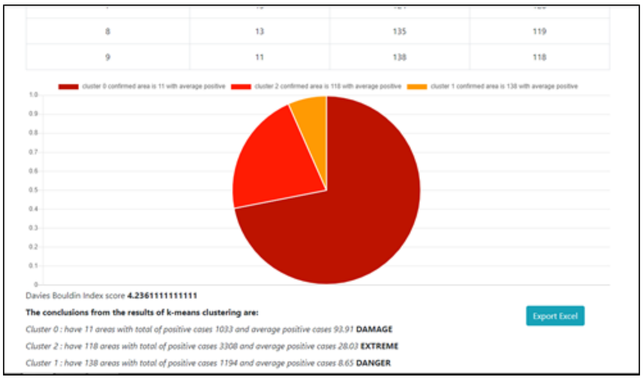

Fig 4. Regional Mapping Based on the Number of COVID-19 Cases

The specifications for each cluster can be defined as consisting of an area with a level of damage based on the average number of positive cases and the total number of positive cases in each cluster. There is an area with 'damage' status consisting of 11 sub-districts with an unlimited number of positive cases of 1,033 and an average number of cases of 93 per sub-district. There is an area with 'extreme' status consisting of 118 sub-districts, with the number of positive instances reaching 3,308, and the average number of cases is 28.

Other areas with 'danger' status consist of 138 sub-districts, with the number of positive cases reaching 1194 and the average number of patients being 8. This area labeling is based on the rule of the average number of positive points that occur in each cluster member. The greater the average number of cases in one cluster, the more it shows the severity of that area in terms of the spread of positive COVID-19 cases in DKI.

To get an overview of areas clustered for severity, the prototype can also define regional specifications based on demographic data in the form of regional density and location. This is important because there are times when the number of positive cases looks normal but occurs in areas with high density. This can be a focus for decision-makers to see whether a positive case has high penetration in dense regions. To visualize the condition of the clustered areas, a feature was developed to view the regional distribution of a formed cluster. The Model Preview feature looks like in Figure 5.

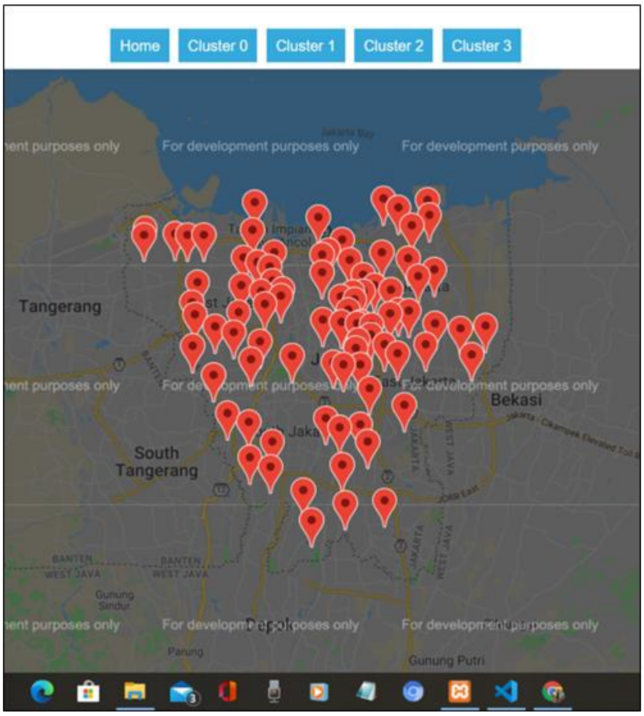

Fig 5. Distribution of positive Covid case points based on clusters

4. Conclusion

The prototype can import data and regional unit data with the number of positive cases at a particular time. The data collection timeline can be set so that users can form their cluster model that will be produced, for example, a timeline at the start of the PSBB period, at a specific period, or other moments. All specified timelines will become header attributes of the dataset, which will be modeled with the k-means algorithm.

The prototype can also determine the number of k desired by the user so that the ideal cluster shape can be tested to get the best parameters as seen from the Davies Bouldin Index (DBI) value. The modeling results have been verified from the centroid values formed so that when the cluster formation iteration stops, it has a final centroid value.

The application prototype is not only able to produce the results of the clusters formed in the form of member specifications for each cluster but also a summary of the statistical processing values in the form of the average number of positive cases from the clusters formed and the total number of issues that occurred in the clusters formed. These results can also be linked to demographic data for each region, such as regional density and area of each cluster member.

Another feature developed from the prototype application is to see the visual distribution of the areas of the formed clusters. In this way, this visualization can also be contrasted in appearance based on the regional density data displayed, in addition to displaying data on the number of positive COVID-19 cases that have occurred. With the ease of selecting a timeline dataset, the current study can be a preliminary study capable of producing a regional clustering model based on the chosen timeline.

The resulting cluster model that is formed will change according to the conditions of the data entered and display different visualization dimensions. However, the limitation of the current study is that it only has quite limited data coverage, in the form of data on the distribution of COVID-19 cases in DKI Jakarta within a period of ninety days since the case was first discovered, and cannot describe the distribution of up-to-date data because various policies have not been implemented. During the data collection period, the data analyzed in this study is limited to data on the distribution of new positive cases in various sub-districts in DKI Jakarta.

References

[1] Kemenkes RI, "Current Situation of the Spread of COVID-19 Data reported until April 2, 2020," *Kementerian Kesehatan Republik Indonesia*, 2020. https://covid19.kemkes.go.id/situasi-infeksi-emerging/info-corona-virus/situasi-terkini-perkembangan-coronavirus-disease-covid-19-3-april-2020/#.Xov_Fy2B01I (accessed Apr. 03, 2020).

[2] Kemenkes RI, "Current Situation of the Spread of COVID-19 Data reported until April 3, 2020," *Kementerian Kesehatan Republik Indonesia*, 2020.

[3] W. I. A. Chafid Nurul, "Implementation of Data Mining for Clustering Areas of Dengue Fever Spread in South Tangerang City Using the K-Means Algorithm," *J. Satya Inform.*, vol. 3, no. 1, pp. 12–24, 2018.

[4] B. M. Metisen and H. L. Sari, "Clustering Analysis Using the K-Means Method in Grouping Product Sales at Fadhila Supermarket," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.

[5] Eric Fammaldo and Lukman Hakim, "Application of the K-Means Clustering Algorithm for Grouping Family Welfare Levels for the Smart Indonesia Card Program," *J. Ilm. Teknol. Inf. Terap.*, vol. V, no. 1, pp. 24–32, 2018.

[6] A. Supriyanto, B. Basukianto, and J. A. Rozaq, "MSME Clustering and Map-Based Regional Potential as a Regional Economic Development Strategy," *J. Pekommas*, vol. 2, no. 2, p. 143, 2017, doi: 10.30818/jpkm.2017.2020204.

[7] A. H. Nasyuha, Z. Zulham, and I. Rusydi, "Implementation of K-means algorithm in data analysis," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 20, no. 2, pp. 307–313, 2022, doi: 10.12928/TELKOMNIKA.v20i2.21986.

[8] I. B. G. Sarasvananda, R. Wardoyo, and A. K. Sari, "The K-Means Clustering Algorithm With Semantic Similarity To Estimate The Cost of Hospitalization,"

IJCCS (Indonesian J. Comput. Cybern. Syst., vol. 13, no. 4, p. 313, 2019, doi: 10.22146/ijccs.45093.

[9] B. Chong, "K-Means Clustering Algorithm: a Brief Review," *Acad. J. Comput. Inf. Sci.*, vol. 4, no. 5, pp. 37–40, 2021, doi: 10.25236/ajcis.2021.040506.

[10] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-Means Clustering Algorithms: A Comprehensive Review, Variants Analysis, And Advances In The Era Of Big Data," *Inf. Sci. an Int. J.*, vol. 622, no. C, 2023.

[11] L. N. C. P. K, G. Suryanarayana, N. Swapna, T. Bhaskar, and A. Kiran, "Optimizing K-Means Clustering using the Artificial Firefly Algorithm," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 9s, pp. 461–468, 2023.

[12] R. Setiawan, "Application of Data Mining Using the K-Means Clustering Algorithm to Determine New Student Promotion Strategies," *J. Lentera ICT*, vol. 3, no. 1, pp. 76–92, 2016.

[13] Jakarta Smart City, "Policy timeline for handling the COVID-19 pandemic in Jakarta," *Jakarta Smart City*, 2020. https://smartcity.jakarta.go.id/blog/507/linimasa-kebijakan-penanganan-pandemi-covid-19-di-jakarta (accessed Jul. 21, 2020).

[14] Kompas.com, "Latest data on Corona cases in RI as of March 6, 2020," *Kompasdotcom*, pp. 3–7, 2020.

[15] Kompas.com, "Anies Issues Gubernatorial Regulation regarding PSBB in DKI Jakarta, Implemented Friday 10 April," *Kompasdotcom*, pp. 7–12, Apr. 10, 2020. [Online]. Available: https://megapolitan.kompas.com/read/2020/04/09/22340801/anies-terbitkan-pergub-soal-psbb-di-dki-jakarta-diterapkan-jumat-10-april?page=all

[16] Kompas.com, "DKI Jakarta PSBB has been extended by 28 days until May 22, 2020," *Kompasdotcom*, pp. 20–25, Apr. 22, 2020. [Online]. Available: https://news.detik.com/berita/d-4987059/psbb-dki-jakarta-diperpanjang-28-hari-hingga-22-mei

[17] AyoJakarta.com, "Anies Extends PSBB Jakarta Again, Starting May 22 to June 4, 2020," *AyoJakartadotcom*, vol. 52, pp. 1–6, 2020. [Online]. Available: https://www.ayojakarta.com/read/2020/05/19/17987/anies-perpanjang-lagi-psbb-jakarta-mulai-22-mei-hingga-4-juni-2020

[18] Pemrov. DKI Jakarta, *Governor's Decree Number 563 of 2020 Regarding the Implementation of Stages and Implementation of PSBB Activities during the Transition Period*, 2020. [Online]. Available: https://covid19.go.id/p/protokol/kepgub-nomor-563-tahun-2020-perihal-pemberlakuan-tahapan-dan-pelaksanaan-kegiatan-aktivitas-psbb-pada-masa-transisi

[19] Dinas Kependudukan DKI Jakarta, "Population Density Data per Subdistrict in 2019," *Dinas Kependudukan DKI Jakarta*, 2019. https://data.jakarta.go.id/dataset/datadkimenurutkepadatanpenduduk/resource/03266a68-1c1b-4848-aebe-f671e4e41ee6