Analysis of The Effectiveness of Covid-19 Prevention Measures in Manado City Using the K-means Method

Analisis Efektivitas Tindakan Pencegahan Covid-19 di Kota Manado Menggunakan Metode K-means

Liza Wikarsa[1)], Steven Pandelaki[2)], Henrico Kurnia[3)]
[1, 2, 3]Universitas Katolik De La Salle Manado
[1, 2, 3]Kairagi I Kombos Manado, Belakang Wenang Permai II, Manado, 95000, Telp: 0431-813148, Fax: 0431-813160

lwikarsa@unikadelasalle.ac.id[1)], spandelaki@unikadelasalle.ac.id[2)], 16013030@unikadelasalle.ac.id[3)]

Diterima : 11 Oktober 2021 || Revisi : 20 Juni 2022 || Disetujui: 21 Juni 2022

Abstract – Covid-19 is an infectious disease caused by the Coronavirus, spreading quickly and resulting in death. Thus, the government of Manado city recommends all the residents always comply with the Covid-19 health protocols. The study aimed to investigate the public's response to the effectiveness of the Covid-19 prevention measures imposed by the local government. The prevention measures include using masks, healthy lifestyles, clean lifestyles, and social distancing. K-means clustering was used to determine the level of effectiveness of prevention measures into four clusters, namely very effective (C1), moderately effective (C2), ineffective (C3), and very weak (C4). The algorithm yields consistent results despite the difficulty of predicting the K-values or the number of clusters at the beginning of the algorithm. As a result, 58% of respondents consider the prevention measures very effective. They have a high level of Covid-19 awareness and the consequences of violating the health protocols, even though about 10% of the respondents doubted the health protocols in Manado were effective due to the high number of death cases daily. The results can be used as insights for the local government to stop the spread of the Coronavirus in Manado city.

Keywords: Covid-19, Effectiveness, K-means, Manado, Health Protocol.

Abstrak *– Covid-19 adalah penyakit menular yang disebabkan oleh Virus Corona. Penyebarannya sangat cepat dan dapat mengakibatkan kematian. Untuk itu, Pemerintah Kota Manado menghimbau kepada seluruh warganya untuk selalu mematuhi protokol kesehatan Covid-19. Tujuan dari penelitian ini adalah untuk mengetahui respon masyarakat terhadap efektivitas tindakan pencegahan Covid-19 yang diberlakukan oleh pemerintah daerah. Langkah pencegahannya adalah penggunaan masker, pola hidup sehat, pola hidup bersih, dan social distancing. K-means clustering digunakan untuk mengetahui tingkat efektivitas tindakan pencegahan menjadi empat cluster, yaitu sangat efektif (C1), cukup efektif (C2), tidak efektif (C3), dan sangat tidak efektif (C4). Algoritma ini menghasilkan hasil yang konsisten meskipun sulit untuk memprediksi nilai-K atau jumlah cluster di awal algoritma. Hasil penelitian menunjukkan bahwa 58% responden menganggap tindakan pencegahan sangat efektif. Mereka memiliki tingkat kesadaran Covid-19 yang tinggi dan konsekuensi melanggar protokol kesehatan. Namun, sekitar 10% responden meragukan protokol kesehatan di Manado efektif karena tingginya angka kematian setiap hari. Hasil kajian tersebut dapat dijadikan sebagai wawasan bagi pemerintah daerah untuk menghentikan penyebaran virus Corona di kota Manado.*

Kata Kunci: *Covid-19, Efektivitas, K-means, Manado, Protokol Kesehatan.*

INTRODUCTION

Covid-19 is an infectious disease caused by the Coronavirus and was first discovered in December 2019 in Wuhan, China (Sabarudin, et al., 2020). The Coronavirus is currently the center of attention worldwide because of its swift spread and can result in death. The World Health Organization (WHO) issued various measures to prevent the spread of the virus. Several policies issued by WHO are always using masks, creating a healthy lifestyle, creating a clean lifestyle, and always keeping a distance (World Health Organization, 2021). These measures are considered adequate by WHO to prevent the spread of the Coronavirus. The Indonesian government has agreed to implement these actions in the lives of the Indonesian people to prevent the spread of the Coronavirus (Sabarudin, et al., 2020).

Manado City is one of the cities in Indonesia that is currently affected by the Corona Virus. The Manado city government recommends all people always

doi: 0.30818/jpkm.2022.2070103

comply with the Covid-19 protocol policies that have been made. The number of positive patients infected in Manado on October 1st, 2021, has reached 34,115, and of that, there were 1,022 deaths (Corona Sulutprov, 2021). Even with the increasing numbers of recovery (32,347 cases), the chain of Coronavirus spread in this city has not entirely stopped. Hence, the government and community must work together hand in hand to increase awareness of the prevention measures and take action against those who ignore the health protocols as they can jeopardize the safety of many people (Yatimah, et al., 2020).

The effectiveness of prevention measures taken by the community is necessary to prevent the spread of this virus. Effectiveness means that an action taken can bring results or affect the fundamental purpose of the action (Kemendikbud, 2021). According to the Ministry of Home Affairs Research and Development, the measurement standard is divided into four levels: very practical, moderately effective, ineffective, and very ineffective (Ayu, 2017). This study will group the community's response to the Covid-19 prevention measures they take and measure how effective these actions are based on the methods used in this study.

K-means method is used in this research to group the level of effectiveness of Covid-19 prevention measures into four clusters (Ayu, 2017), namely very effective (C1), moderately effective (C2), and ineffective (C3), and very ineffective (C4). This unsupervised method will group data into several clusters wherein the data in a group has the same characteristics but differs from other groups (Arsoni, Fitri, & Prasetyo, 2018). They also pointed out that K-means has the advantage of high data size accuracy, so it is relatively more scalable and efficient in processing large amounts of data. However, K-means requires precise numbers in determining the initial centroid because sometimes the cluster results can change so that the grouping becomes unstable (Priati & Fauzi, 2017).

The previous study conducted by Hariyanto and Shita (Hariyanto & Shita, 2018), grouped the potential spread of dengue fever in the city of South Tangerang into three sporadic clusters, potential clusters, and endemic clusters. The results revealed sporadic clusters of 19 villages, potential clusters of 13 villages, and endemic clusters of 20 villages from a total of 52 urban village data. However, this study could not randomly determine the center value or initial centroid, so the results obtained were less than optimal. Another research was conducted by Maulida that grouped tourist visits to top tourist attractions in the province of DKI Jakarta (Maulida, 2018). The data contained eight tourist attractions and the average number of visits to these tourist attractions. This study used three clusters: "C1 = the number of high tourist visits, C2 = the number of tourists visiting medium, and C3 = the number of tourist visits is low" (pp.167). Solichin and Khairunnisa grouped the spread of the Coronavirus (Covid-19) in 44 districts in DKI Jakarta based on the number of ODP (People Under Monitoring), PDP (Patients Under Monitoring), positive cases, patients recovering, and patients dying (Solichin & Khairunnisa, 2020). The results revealed that Cluster 1 obtained 19 districts, Cluster 2 got 23 districts, and Cluster 3 had two districts. System users can determine the number of clusters, but there is no naming of cluster categories to represent a particular situation or purpose.

This study will investigate the effectiveness of the Covid-19 prevention measures in Manado city. The independent variables are the use of masks, a clean lifestyle, a healthy lifestyle, and social distancing. Each of these variables has several supporting indicators. For example, the mask variable has indicators of the type of mask used, how to use it, and how to care for the mask. The contribution of this study is to help the Manado city government to get insights into the public's response to the effectiveness of the Covid-19 prevention measures that have been taken to date.

The organization of this paper is divided into four parts: introduction, research methodology, results and discussion, and conclusions and recommendations. First, the introduction will address the research background to understand better the context in which this study was conducted. Next, it will provide meticulous details of the research methodology, including the software development life cycle. Then, the results and discussion are provided before the conclusion and future works.

RESEARCH METHOD

This study developed a research framework consisting of 4 stages: data collection, pre-processing, processing, and validation, as seen in Figure 1.

Stage I: Data Collection

Currently, the population of Manado city is as much as 527,007 inhabitants (Pemerintah Kota Manado, 2021). Therefore, based on Isaac and Michael's table

with a significance level of 0.05, the minimum number of samples used is 348 people registered with Manado ID cards. In this data collection stage, questionnaires were distributed twice to the respondents with different intents. The first questionnaire was used to find the critical value of each variable of Covid-19 prevention measures. This questionnaire was distributed to 50 people located in Manado city. Meanwhile. For this purpose, 348 residents in Manado were used as the research sample.

The score of each item obtained from the second questionnaire will be added with the results of the importance of the first questionnaire, aiming to obtain the value of each variable which will later be used in implementing the K-means method.

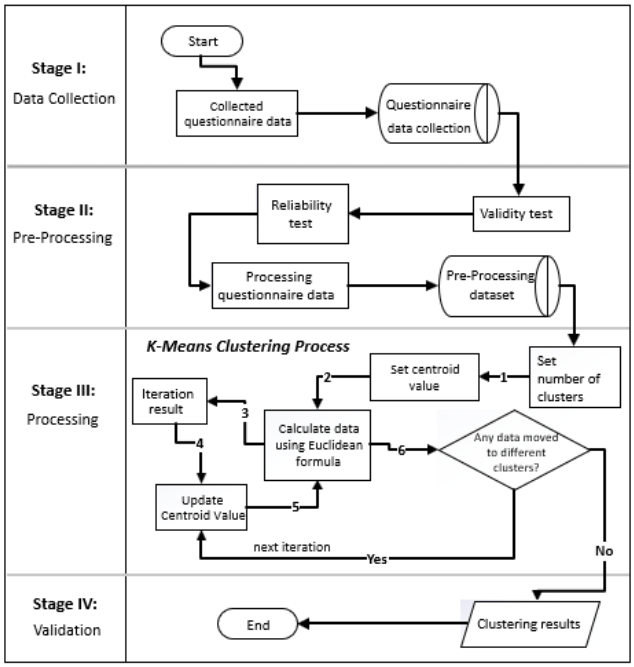


Figure 1 Research Framework

Stage II: Pre-processing

At this stage, the questionnaire data collection will pass several statistical tests, namely the validity and correlation of the data using the Pearson formula, reliability testing using the Cronbach's Alpha formula, and the T-test. These tests were carried out using the SPSS 25 (Statistical Product and Service Solutions) program to see if questionnaire data is feasible to use or not. Data that has passed the statistical test will be processed to be used at the processing stage.

Stage III: Processing

The data collection that has passed the pre-processing stage will be processed using the K-means clustering method. Then, the grouping process will be carried out through the application to be built.

The following will enlist all the steps to implement the K-means method (Arsoni, Fitri, & Prasetyo, 2018; Maulida, 2018):

1. Determine the number of clusters.
2. Determine the initial centroid/center value randomly.
3. Calculate the distance of each data based on the closest distance between the data and the centroid value using the Euclidean Distance formula.

$$d(\mathbf{i},\mathbf{k}) = \sqrt{\sum_{i}^{m}(C_ij - C_kj)^2} \quad (1)$$

Note:
d = distance between x and y
i = data on attribute
k = cluster center data
i = each data
m = number of data
C_ij = data on each data k i
C_kj = data at the center of the cluster to i

4. Group each data based on the closest distance to the centroid value.

$$\min \sum_{k}^{i} -a_{ik}- = \sqrt{\sum_{i}^{m}(C_ij - C_kj)^2} \quad (2)$$

5. Define a new centroid/center value.

$$C_{kj} = \frac{\sum_{k}^{i} x_{ij}}{p} \quad (3)$$

6. Recalculate by repeating steps 3 and 4.
7. If no data has moved to other clusters/groups, the grouping has been declared complete. If there is still data moving, repeat step 3 until the results of the grouping data are the same as the previous iteration.

Stage: Validation

In the last stage, this study will display the results of the clustering in four different levels such as very effective (C1), moderately effective (C2), ineffective (C3), and very ineffective (C4). In addition, each cluster will show how many people are in the cluster in several visual displays such as tables and charts.

Analysis

This study adopted a system development life cycle with four stages: analysis, design, implementation, and testing. Each stage would be addressed according to the research framework that consists of data collection, pre-processing, processing, and validation.

Data Collection

Data on the importance of Covid-19 prevention measures were collected to get a percentage of what actions were considered necessary by the respondents. The first data collection will be a percentage (%) so that it can be multiplied by the data on Covid-19 prevention measures taken by the community to find the score for each variable. It was noted that the first questionnaires were distributed as a pilot study to try out the study with fewer respondents. The respondents could give honest feedback on the questionnaire design. Also, it allowed us to adjust the main study. The result of this first data collection was the independent variables for this research.

Tabel 1 Research Variables and Indicators

Variable	Indicator	Initial (%)	Final (%)	Variable Weight
Q1: Mask (Atmojo, 2020)	Q1.1: Type of masks	94%	9.71%	27%
	Q1.2: How to use	90%	9.30%	
	Q1.3: Maintenance	76%	7.85%	
Q2: Clean Lifestyle (Nakoe, Lalu, & Mohamad, 2020)	Q2.1: Wash hands with soap	92%	9.50%	25%
	Q2.2: Use of hand-sanitizer	82%	8.47%	
	Q2.3: Use of disinfectant	70%	7.23%	
Q3: Healthy Lifestyle (Asri, Lestarini, Husni, Muspita, & Hadi, 2021)	Q3.1: Do exercise	86%	8.88%	24%
	Q3.2: Take vitamin	66%	6.82%	
	Q3.3: Consuming nutritious food and drinks	80%	8.26%	
Q4:Social Distancing (Mona, 2020)	Q4.1: Keep a minimum distance of 1 meter	94%	9.71%	24%
	Q4.2: Awareness of the environment	76%	7.85%	
	Q4.3: Self-isolation/stay at home	62%	6.40%	

The Table 1 is the result of the percentage of importance of each variable supporting indicators given by 50 respondents through a Google form questionnaire.

Once all the independent variables were identified in the first questionnaire, the second questionnaire was distributed to 348 Manado residents. This second data collection was the adjusted questionnaires to determine the effectiveness of Covid-19 prevention measures in this city based on the independent variables identified from the first questionnaire. Again, the Likert scale was used to score the answers for each question in both questionnaires.

Pre-Processing

In this stage, the questionnaires were tested for their validity, reliability, and T-test. The validity and correlation test can be seen in Table 2 below.

Table 2 Questionnaire Validity and Correlation Test

Item Code	r_{hitung}	r_{tabel}	Significance	Validity
Q1.1	0.626**	0.138	.000	Valid
Q1.2	0.756**	0.138	.000	Valid
Q1.3	0.796**	0.138	.000	Valid
Q2.1	0.713**	0.138	.000	Valid
Q2.2	0.765**	0.138	.000	Valid
Q2.3	0.726**	0.138	.000	Valid
Q3.1	0.510**	0.138	.000	Valid
Q3.2	0.737**	0.138	.000	Valid
Q3.3	0.361**	0.138	.000	Valid
Q4.1	0.709**	0.138	.000	Valid
Q4.2	0.636**	0.138	.000	Valid
Q4.3	0.560**	0.138	.000	Valid

The data will be considered valid if the value of the r-count exceeds the number of r-table. The value of the r-table itself is determined based on the number of respondents. The data validity was tested using Pearson's formula, and the results showed that all questionnaire items used were proven valid or could be used. Table 2 shows that the Pearson's Correlation value on each connected variable has a significance value of 0.000 < 0.138, so it can be concluded that each variable tested has a correlation (Zahra & Rina, 2018).

The reliability test for the research instrument is shown in Table 3.

Tabel 3 Reliability Test

Chronbach's Alpha	Number of Items
0.874	12

Based on the tests, it was found that Chronbach's alpha value of 0.874 which if the alpha value obtained is 0.6 or more, the questionnaire can be declared reliable (Zahra & Rina, 2018).

The T-test is one of the research hypotheses tests that aims to determine whether the independent

variable partially affects the dependent variable (Zahra & Rina, 2018).

Tabel 4 T-test

Kode item	T_{hasil}	T_{tabel} 0.05	Acceptance
Q1.1	4.725	1.646	Accepted
Q1.2	4.042	1.646	Accepted
Q1.3	5.251	1.646	Accepted
Q2.1	5.791	1.646	Accepted
Q2.2	3.871	1.646	Accepted
Q2.3	3.751	1.646	Accepted
Q3.1	5.750	1.646	Accepted
Q3.2	5.201	1.646	Accepted
Q3.3	3.984	1.646	Accepted
Q4.1	3.878	1.646	Accepted
Q4.2	3.903	1.646	Accepted
Q4.3	3.641	1.646	Accepted

Based on the tests carried out on each item, the results obtained t exceeds the t-table so that it can be concluded that the hypothesis is accepted.

Processing

The following is an example of data that has been processed by multiplying the indicator of the importance of the action (%) with the result of the indicator of the action taken by one respondent.

1. Scores were obtained from collecting questionnaires about Covid-19 prevention measures taken by the community.

Table 5 Scores for Questionnaires

Question	Score	Question	Score
Q1.1	4	**Q3.1**	5
Q1.2	4	**Q3.2**	1
Q1.3	3	**Q3.3**	5
Q2.1	3	**Q4.1**	4
Q2.2	3	**Q4.2**	4
Q2.3	1	**Q4.3**	1

2. The score of the actions taken by the respondent is multiplied by the level of importance of the actions in Table 1 so that the results are as follows.

Table 6 Percentages for Indicators Scoring

	Score	Result		Score	Result
Q1.1	4 x 9.71%	0.39	**Q3.1**	5 x 8.88%	0.44
Q1.2	4 x 9.30%	0.37	**Q3.2**	1 x 6.82%	0.07
Q1.3	3 x 7.85%	0.24	**Q3.3**	5 x 8.26%	0.41
Q2.1	3 x 9.50%	0.29	**Q4.1**	4 x 9.71%	0.39
Q2.2	3 x 8.47%	0.25	**Q4.2**	4 x 7.85%	0.31
Q2.3	1 x 7.23%	0.07	**Q4.3**	1 x 6.40%	0.06

3. After the results are obtained, each item belonging to the same variable or one type of indicator will be added up and divided by three so that the average value is obtained as follows.

Table 7 Percentages for Indicators Scoring

Mask (Q1)	Clean Lifestyle (Q2)	Healthy Lifestyle (Q3)	*Social Distancing* (Q4)
0.33	0.20	0.31	0.25

4. Data processing will be carried out on all data totaling 348 samples. After all the data has been processed, the data will be saved with the CSV extension so that it can be entered into the application to be built.

Table 8 Example of the Processed Data for 1 Respondent

Id Respondent	001
Age	17-24
Gender	Man
Mask (Q1)	0.331956
Clean Lifestyle (Q2)	1.357438
Healthy Lifestyle (Q3)	0.30854
Social Distancing (Q4)	0.25551

5. The implementation of K-means followed the steps:
   a. Determine the number of clusters.
      There are 4 clusters such as very effective (C1), moderately effective (C2), ineffective (C3), and very ineffective (C4).
   b. Determine the initial centroid/center value randomly.

Table 9 Initial Centroid

Cluster	Mask	Clean Lifestyle	Healthy Lifestyle	Social Distancing
C1	0.30096	2.76309	0.353994	0.347107
C2	0.3333	2.38154	0.265152	0.21832
C3	0.24242	1.69077	0.251377	0.26584
C4	0.08953	1.0241	0.0798898	0.0798898

   c. Calculate the distance of each data based on the closest distance between the data and the centroid value using the Euclidean Distance formula.
   d. Group each data based on the closest distance to the centroid value.
   e. Define a new centroid/center value.
   f. Recalculate by repeating steps C to D.
   g. If no data has moved clusters/groups, the grouping has been declared complete. If there is still data moving, repeat step 3 until the results of the grouping data are the same as the previous iteration.

Table 10 New Centroid Values for the Last Iteration

Cluster	Mask	Clean Lifestyle	Healthy Lifestyle	Social Distancing
C1	0.388345	3.04146	0.312489	0.322161
C2	0.342002	2.2579	0.27382	0.287708
C3	0.279104	1.51005	0.255076	0.241276
C4	0.15404	0.83892	0.123661	0.108892

The findings showed that no data moved to another cluster after the 5th iteration. The final result presents the respondents' responses regarding the effectiveness of the Covid-19 prevention measures that have been taken to date are provided in Table 11.

Table 11 Final Results of Clustering

R*	C1	C2	C3	C4	Result
1	1.68629	0.901758	0.170716	0.57871	3
2	0.30663	0.485018	1.23551	1.92387	1
3	1.6899	0.905102	0.169463	0.57204	3
4	0.65078	0.17207	0.897218	1.58211	2
5	0.97189	0.202623	0.576284	1.26874	2

*R=respondent

Requirements Specification

1. The application can upload respondent data that has been processed through the pre-processing stage into the application as a CSV file.
2. The application can display the respondent data on the home page.
3. The initial centroid of the application can be processed randomly by taking respondent data that has been previously labeled in the pre-processing stage.
4. The application can perform processing by calculating the Euclidean distance of respondent data with the initial centroid, which has been processed randomly.
5. The new centroid from each subsequent iteration can be updated automatically using the K-means Clustering method.
6. The application can display the entire list of iterations from the grouping results so that the user can select an iteration to see the details of the iteration results.
7. The results of each iteration can be displayed in the form of a chart and several tables, such as the number of respondents belonging to each cluster, the new centroid used in the iteration, and the results of the calculation of the Euclidean distance.

Validation

The 5-Fold cross-validation (K=5) was used in this research by splitting the dataset into five folds. The 5-fold cross-validation was used to train five different models to understand better what was going on. In addition, the trained five models helped measure the accuracy of those models. In the 1st iteration, the 1st fold was used to test the classifier, and the rest of the dataset was used to train the classifier. In the 2nd iteration, the 2nd fold was used as the testing set while the rest of the datasets served as the training set. This process was repeated until each fold of the 5-fold cross-validation had been used as the testing set.

Design

This section will present a flowchart demonstrating the implementation of the K-Mean method in the application, storyboard design, and database design.

1. First, the user must upload a CSV file containing the dataset.
2. Define the number of clusters *K*.
3. Initialize centroids by shuffling the dataset and randomly selecting *K* data points for the centroids.
4. Apply the Euclidean formula to the data by which it calculated the sum of the squared distance between data points and all centroids.
5. Assign each data point to the closest cluster (centroid), whereby a minimum distance will group the data.

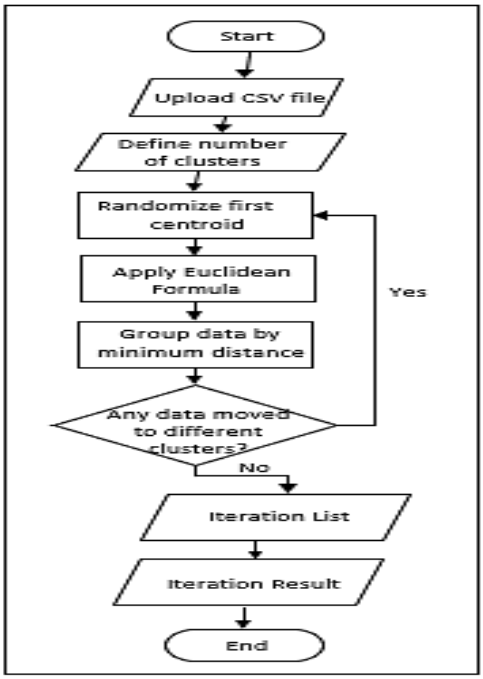


Figure 2 K-means Method

6. Compute the centroids for the clusters by taking the average of all the data points belonging to each cluster.
7. Repeating the steps until there is no change to the centroids, whereby the assignment of data points to clusters is no longer changing.
8. The application will present the data computed in each iteration until the last iteration.

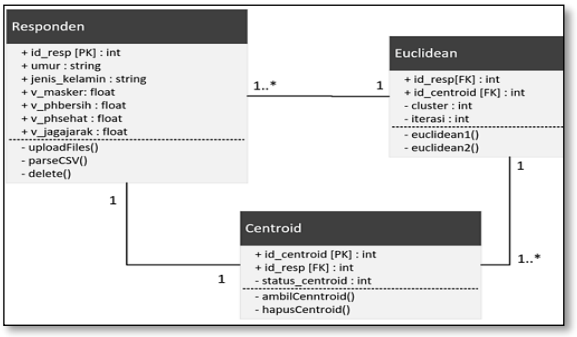


Figure 3 Database Design

Figure 3 presents the database design for the application to contain the values of the responses gathered from the questionnaires, the values of centroids after each iteration, and the values of the sum of the squared distance between data points and all centroids using the Euclidean formula.

Data Collection

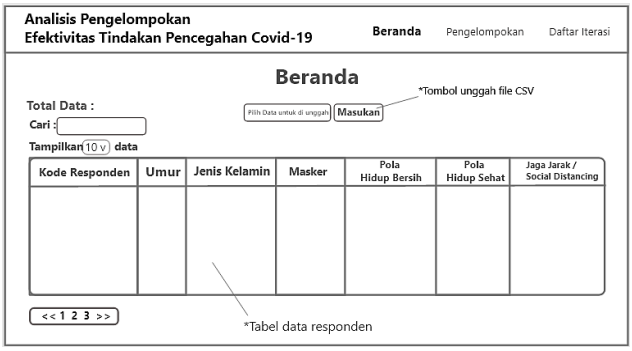


Figure 4 A Page for Uploading CSV File

On this page, the user must upload the CSV file. The CSV file has the responses for the independent values gathered from the distributed questions to 348 respondents.

Pre-processing

Once the user has uploaded the CSV file, the content of this file will be displayed on this page. The data in this file will be later processed on this page by pressing "Pengelompokan," located on the top menu.

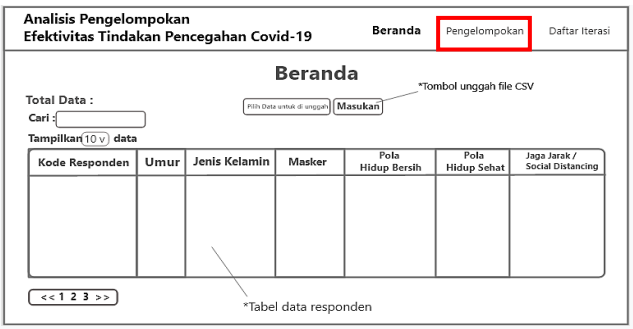


Figure 5 Pre-processing the Data

Processing

One can press the "Acak Centroid" button to randomize the initial centroid. This application will display the independent values, namely mask, clean lifestyle, healthy lifestyle, and social distancing, along with its associated clusters, as shown in Figure 6.

Figure 6 Randomized the Initial Centroid

To start the grouping process, the user can press the "Calculate" button and direct it to the Iteration List in the following figure.

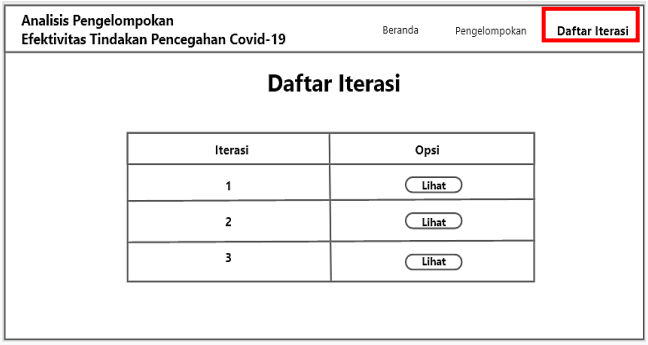


Figure 7 Number of Iterations

Validation

This part provides the results, as seen in Figure 8.

Implementation

Table 12 shows the development environment consisting of the specifications of hardware and software required for the application development.

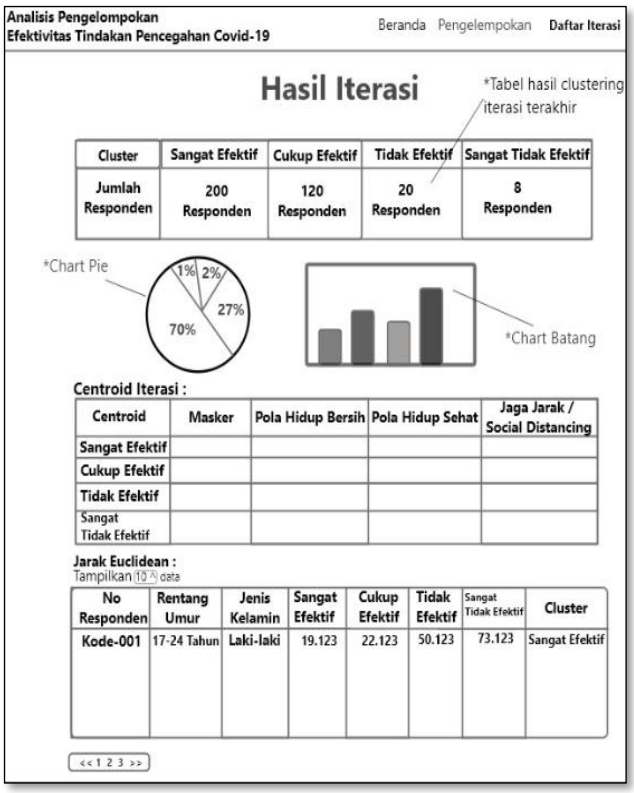


Figure 8 The Final Results

Table 12 Development Environment

Sumber Daya	
Hardware	**Laptop's specifications:** CPU Intel Core i58250U, SSD 256 GB, RAM 8GB, Windows 10.
Software	1. Python 3.9 for the code 2. Python's libraries 3. phpMyAdmin for DBMS 4. Visual Studio Code for writing the codes. 5. Mozilla Firefox is the selected browser to display the application contents.

Data Collection

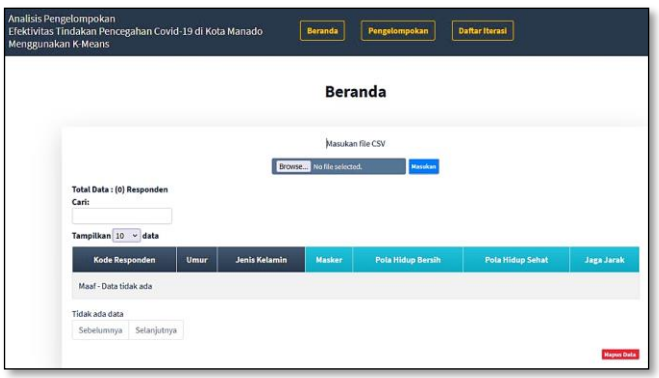


Figure 9 Uploading CSV File

The CSV file contains the data gathered from the questionnaires.

Pre-processing

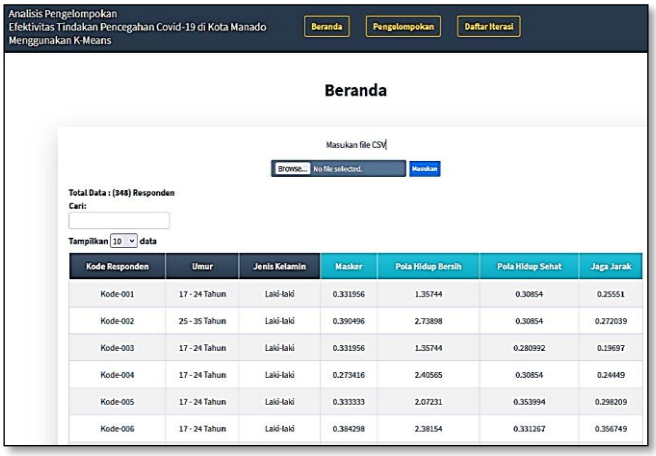


Figure 10 Pre-processed Data

Figure 10 shows the data from the CSV file, which includes the respondent codes, age, gender, and the values for mask, clean lifestyle, healthy lifestyle, and social distancing.

Processing

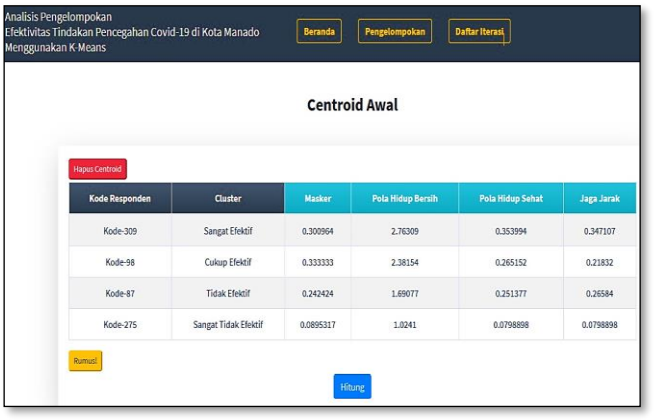
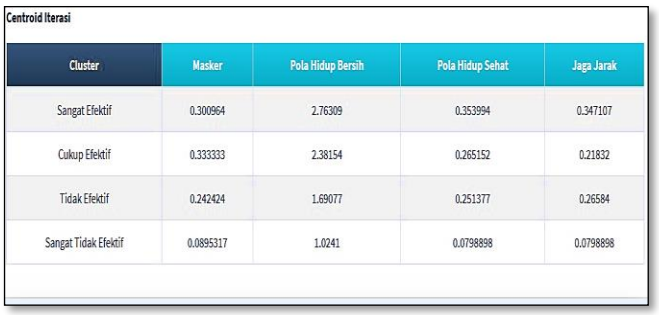


Figure 11 Initial Centroid

Figure 11 displays the values of the initial centroid that are required when implementing the K-means method.

Figure 12 The number of iterations

Figure 12 shows how many iterations were done concerning the uploaded data. This application allows the user to select any of the iterations in the table to view.



Centroid Iterasi

Cluster	Masker	Pola Hidup Bersih	Pola Hidup Sehat	Jaga Jarak
Sangat Efektif	0.300964	2.76309	0.353994	0.347107
Cukup Efektif	0.333333	2.38154	0.265152	0.21832
Tidak Efektif	0.242424	1.69077	0.251377	0.26584
Sangat Tidak Efektif	0.0895317	1.0241	0.0798898	0.0798898

Figure 13 The Values of Initial Centroid for 1st Iteration

Rincian Iterasi ke - 1

Jumlah Data Cluster:

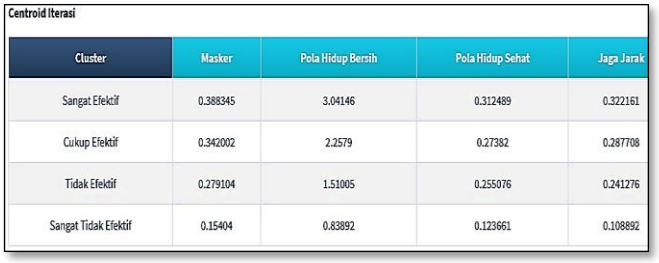
Total Responden	Sangat Efektif	Cukup Efektif	Tidak Efektif	Sangat Tidak Efektif
348	203	107	29	9

Pie Chart | Bar Chart

Figure 14 The result for 1st iteration

Figure 13 and Figure 14 present the values of initial centroids and the result of 1st iteration after calculating the distance of each data based on the closest distance between the data and the centroid value using the Euclidean Distance formula.

Figure 15 and Figure 16 exhibit the last iteration. It provides the centroid values and the result of calculating the data distance using the Euclidean Distance formula.



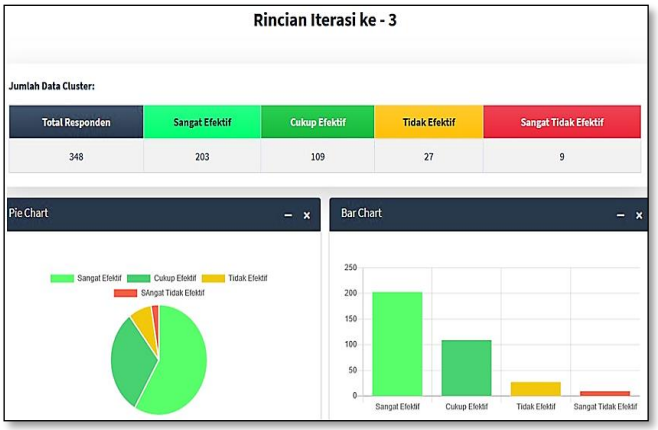
Centroid Iterasi

Cluster	Masker	Pola Hidup Bersih	Pola Hidup Sehat	Jaga Jarak
Sangat Efektif	0.388345	3.04146	0.312489	0.322161
Cukup Efektif	0.342002	2.2579	0.27382	0.287708
Tidak Efektif	0.279104	1.51005	0.255076	0.241276
Sangat Tidak Efektif	0.15404	0.83892	0.123661	0.108892

Figure 15 The Centroid Values for 3rd Iteration

Validation

After applying the K-means method, it can finally determine which cluster each respondent belongs to, as shown in Figure 17.



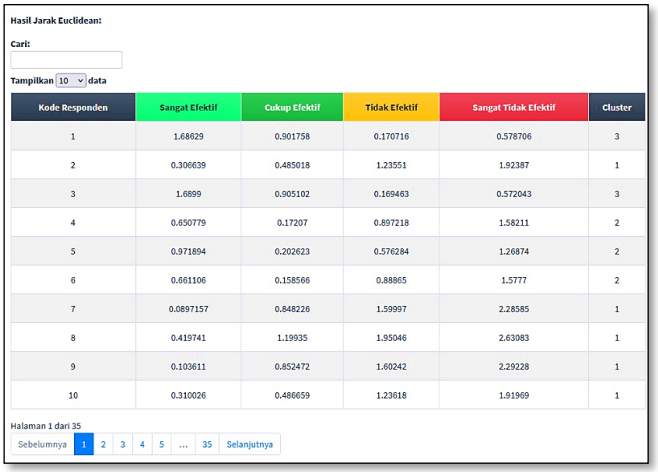
Rincian Iterasi ke - 3

Jumlah Data Cluster:

Total Responden	Sangat Efektif	Cukup Efektif	Tidak Efektif	Sangat Tidak Efektif
348	203	109	27	9

Pie Chart | Bar Chart

Figure 16 The Final Result for 3rd iteration



Hasil Jarak Euclidean:

Cari:

Tampilkan 10 data

Kode Responden	Sangat Efektif	Cukup Efektif	Tidak Efektif	Sangat Tidak Efektif	Cluster
1	1.68629	0.901758	0.170716	0.578706	3
2	0.306639	0.485018	1.23551	1.92387	1
3	1.6899	0.905102	0.169463	0.572043	3
4	0.650779	0.17207	0.897218	1.58211	2
5	0.971894	0.202623	0.576284	1.26874	2
6	0.661106	0.158566	0.88865	1.5777	2
7	0.0897157	0.848226	1.59997	2.28585	1
8	0.419741	1.19935	1.95046	2.63083	1
9	0.103611	0.852472	1.60242	2.29228	1
10	0.310026	0.486659	1.23618	1.91969	1

Halaman 1 dari 35

Sebelumnya 1 2 3 4 5 ... 35 Selanjutnya

Figure 17 Clustering for Each Respondent

Testing

The 5-Fold cross-validation (K=5) was used in this research by splitting the dataset into five folds. Each validation had different initial centroids to generate the grouping results. The result is shown in Table 13 below.

Table 13. The Results of 5-Fold Cross-Validation

Testing	C1	C2	C3	C4
1	203	107	29	9
2	200	108	28	12
3	203	109	27	9
4	199	115	23	11
5	203	106	27	12

Based on the results above, the range value of this validation is between 12 and 203. These range values are divided into four ranges: 1) 12-51, 2) 51-102, c) 103-153, and d) more than 154 and less than 204. Thus, Table 13 shows that the prevention measures taken by the community are regarded to be very effective in preventing the spread of this virus, as the values in C1 are above 154. The use of masks, a clean lifestyle, a healthy lifestyle, and social distancing have been applied daily since the Coronavirus outbreak in Manado city. Despite that, a small group of respondents

considered the prevention measures very ineffective due to the number of people infected by this virus.

RESULTS AND DISCUSSION

The K-means method is relatively simple to implement. For K-means clustering to be effective, the number of clusters *K* must be specified at the beginning of the algorithm. In the processing stage, four clusters were defined, namely very effective (C1), moderately effective (C2), ineffective (C3), and very ineffective (C4). The algorithm performed shuffling on the dataset and thus randomly selected *K* data points for the initial centroids. It was noticeable that when the numbers of data were not so many, the initial grouping would affect the cluster significantly. It was also challenging to predict the K value and know the actual cluster. Using the same data, should it be inputted in a different order, it might generate a different cluster when the number of data is a few. In other words, the K-means method is sensitive to the initial condition as it strongly impacts the results. Equally important, K-means has trouble clustering where clusters have varied sizes and densities.

In the validation stage, the 5-Fold cross-validation (K=5) was done to check the consistency of the results. K-means clustering often produces varying results on different runs of the algorithm. A random choice of cluster patterns generates different clustering results that can cause inconsistency. This study used the randomized values for the initial centroids used for the 5-fold cross-validation, and the results show consistency despite different initial values of the centroids. In this study, more than half of the respondents, 203 out of 348 (58%), believe that the prevention measures for Covid-19 in Manado city are very effective, as shown in Figure 16. These respondents abide by the health protocols imposed by the government for the safety of those who live in Manado. They are also aware of the consequences of not taking the health protocols seriously, as livelihood is at risk. On the contrary, there is still a small group of respondents (10%) from C3 and C4 who consider the health protocols in Manado to be infective since the death cases caused by this virus are relatively high, with approximately 100-200 deaths daily (Informasi dan Koordinasi Covid-19 Pemerintah Kota Manado, 2021). The score is decreasing from C1 to C4 because the level of awareness and consciousness towards the health protocols of Covid-19 in this city was regarded as high due to an intense fear of death.

Based on the test results, it can be concluded that the implementation of the K-means method to determine the effectiveness of the prevention measures for Covid-19 has been successful in providing consistent results despite the difficulty of predicting the K-values or the number of clusters at the beginning of the algorithm.

CONCLUSION AND RECOMMENDATIONS

The following conclusions can be made after conducting a thorough study on the effectiveness of prevention measures for Covid-19 in Manado city:

1. The initial centroid dramatically impacts the final result of the clustering.
2. Applications can group the effectiveness of Covid-19 prevention measures in Manado City using K-means into four clusters: very practical, moderately effective, ineffective, and very ineffective.
3. Based on the tests, the application can be said to be reliable due to its ability to produce consistent clustering results.
4. An application was created to assist the Manado city government in evaluating the level of interest and actions taken by the community to stop the current spread of the Coronavirus.

Recommendations for future works are as follows:

1. This research can further be developed by comparing the K-means Clustering method with other Clustering methods to expand the topic research.
2. For the application development, it is expected to display groupings based on the effectiveness of actions from specific sub-topics, for example, the effectiveness of using masks and the effectiveness of distancing measures.
3. Adding more visualization features to view graph plots that can spread according to the distance between the data and the centroid is highly recommended for application development.

ACKNOWLEDGMENT

We want to convey our gratitude for the support given by Universitas Katolik De La Salle Manado. Also, we thanked the respondents for their participation in this research.

REFERENCES

Arsoni, Fitri, H., & Prasetyo, E. (2018). Penerapan Metode Clustering dengan Algoritma K-means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi

Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *Semesta Teknika*, *21*, 60-64.

Asri, I. H., Lestarini, Y., Husni, M., Muspita, Z., & Hadi, Y. A. (2021). Edukasi Pola Hidup Sehat di Masa Covid-19. *Abdi Populika*, *2*, 56-63.

Atmojo, J. T. (2020). Penggunaan Masker Dalam Pencegahan Dan Penanganan Covid-19: Rasionalitas, Efektivitas, dan Isu Terkini. *Journal of Health Research*, *3*,84-92.

Ayu, P. D. (2017). Analisis Pengukuran Tingkat Efektivitas dan Efisiensi Sistem Informasi Manajemen Surat STIKOM Bali. *Jurnal Sistem Dan Informatika*, *11*,99-109.

Corona Sulutprov. (2021, October 2nd). Retrieved from Sulutprov: https://corona.sulutprov.go.id/

Hariyanto , M., & Shita, R. T. (2018). Clustering Pada Data Mining Untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-means dan Metode Perhitungan Jarak Euclidean Distance. *SKANIKA*, *1,117*-122.

Informasi dan Koordinasi Covid-19 Pemerintah Kota Manado. (2021, October 11th). Retrieved from Pemerintah Kota Manado: https://covid19.manadokota.go.id/

Kemendikbud. (2021, September 30th). Retrieved from KBBI Daring: https://kbbi.kemdikbud.go.id/entri/efektif

Maulida, L. (2018). Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Provinsi DKI Jakarta. *Jurnal Informatika Sunan Kalijaga*, *2*,167-174.

Mona, N. (2020). Konsep Isolasi Dalam Jaringan Sosial Untuk Meminimalisasi Efek Contagious (Kasus Penyebaran Virus Corona Di Indonesia). *Jurnal Sosial Humaniora Terapan*, *2*,117-125.

Nakoe, M. R., Lalu, N. A., & Mohamad, Y. A. (2020). Perbedaan Efektivitas Hand Sanitizer Dengan Cuci Tangan Menggunakan Sabun Sebagai Bentuk Pencegahan Covid-19. *Jambura Journal of Health Science and Research*, *2*, 65-68.

Pemerintah Kota Manado. (2021, September 30th). Retrieved from Jumlah Pendudukan Kota Manado: https://www.manadokota.go.id/index.php?/Infopenting/detail/jumlah-penduduk-kota-manado

Priati, & Fauzi, A. (2017). Data Mining dengan Teknik Clustering Menggunakan Algoritma K-means pada Data Transaksi Superstore," dalam Seminar Nasional Informatika dan Aplikasinya. *Seminar Nasional Informatika dan Aplikasinya (SNIA) 2017* (pp. 15-19). Cimahi: Universitas Jenderal Achmad Yani.

Sabarudin, Mahmudah, R., Ruslin, Aba, L., Nggawu, L., Syahbudin, Nirmala, F., Saputri, A.I., Hasyim, M. S. (2020). Efektivitas Pemberian Edukasi secara Online melalui Media Video dan Leaflet terhadap Tingkat Pengetahuan Pencegahan Covid-19 di Kota Baubau. *Jurnal Farmasi Galenika*, *6*, 309 –318.

Solichin, A., & Khairunnisa, K. (2020). Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-means. *Fountain of Informatics Journal*, *5*, 52-59.

World Health Organization. (2021, February 25th). *Protect yourself and others from COVID-19*. Retrieved from World Health Organization: http://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public

Yatimah, D., Kustandi, C., Maulidina, A., Irnawan, F., & Andinnari, S. (2020). Peningkatan Kesadaran Masyarakat tentang Pencegahan COVID-19 berbasis Keluarga dengan Memanfaatkan Motion Grafis di Jakarta Timur. *Jurnal Karya Abdi*, 246-255.

Zahra , R. K., & Rina, N. (2018). Pengaruh Celebrity Endorser Hamidah Rachmayanti Terhadap Keputusan Pembelian Produk Online Shop Mayoutfit Di Kota Bandung. *Jurnal Lontar*, 43-57.

Halaman ini sengaja dikosongkan