Determination of Traveler Permission in Precaution Covid-19 Area Based on Precaution Covid-19 Pandemic (PCP) Level

Yeffry Handoko Putra[*1], Lia Warlina[2], Selfa Septiani Aulia[3], Wantoro[4], Dina Fatimah[5]
[1,2,3,4,5]Universitas Komputer Indonesia, Bandung, Indonesia
E-mail: [*1]**yeffryhandoko@email.unikom.ac.id**, [2]lia.warlina@email.unikom.ac.id,
[3]selfa@email.unikom.ac.id, [4]wantoro@email.unikom.ac.id, [5]dina.fatimah@email.unikom.ac.id

Abstract

Indonesia from 2020 to 2021 were exposed to COVID-19 pandemic. Both countries implemented a policy of restricting entry areas based on almost the same criteria, In Indonesia namely as PPKM which applying some level of exposure to those infected with covid-19. The determination of this level was all based on the growth in numbers exposed to covid-19, but on pandemic cases, the number of people who do not suffer from COVID-19 disease but have the same symptoms as the symptoms of COVID-19 also need to be considered as the pandemic agent to their environment. We named it as Precaution Covid-19 Pandemic (PCP) Level. The current level of the COVID-19 pandemic has not been fully determined by this idea. So, the idea of this research is to determine the pre-pandemic or precaution level of covid-19 in an area interfere by surrounding area. PCP level was not based on the growth of those infected with the covid-19 disease, but influenced by the number of patients whose have the symptoms similar to the dominant symptoms of the covid-19. The PCP Level determination can be used for precaution policy and support the previous Level Pandemic Methods. To accomplish this idea, two algorithms are used, they are K-Mean algorithm as a pattern clustering and the AHP algorithm as a level determination of the Covid-19 pandemic. Data are collected from 11 health centers in Bandung Regency. The last thing is two determine the allowance for travel to other town is done by check the tendency for traveler to be infected determine by the PCP Level and using Naïve bayes algorithm which recognizing the pattern of symptoms. The results of this study show that the combination of the three proposed algorithms can be used as alternative to give warrant decision about getting infection of Covid-19

Keywords*—Covid-19, Pandemic, Precaution Level, K-Mean, AHP*

1. INTRODUCTION

The case of the spread of Covid-19 has begun in the Asian continent since 2020, in Indonesia itself Covid-19 entered in Mart 2020. The government has locked down policy (namely PPKM) several areas based on the number of people infected and the speed of growth of Covid-19 patients [1]. At first glance this decision is an normal decision for the case to mitigate the spread, but it is not always the right way. Some adverse effect are happened following the enforcement of PPKM like the reduce of economy activities, the sliding down the take home pay of non sectoral employee. So the idea change to count the prevention beside the mitigation. Prevention is more appropriate to be done in Covid-19 condition when the

vaccination of Covid-19 has been done by almost of people in Indonesia in the end of 2021. The prevention can be done when we know the symptom of Covid-19, however the main problem is no reference or academic researchers has certainty declare the true symptom [2] of Covid-19. At least in this study the dominant symptoms [3] of COVID-19 patients are found by doing survey to appropriate stakeholder. But the problem is that the symptoms of COVID-19 still resemble the common cold. But still for reasons of prevention, it is necessary to choose what is closely-dominant symptoms of Covid-19 patients so that the government can do the prevention action. We underlined that prevention is mean the Covid-19 is not yet happened. The results of the number of patients with symptoms that resemble COVID-19 are named the precaution covid-19 pandemic (PCP) level. Of course, it is not certain that the number of patients with symptoms that resemble COVID-19 will definitely become Covid-19 patients, but the PCP level can at least be used to make strategic preventive actions.

Several studies have used the K-mean algorithm to classify data on COVID-19 patients in Indonesia, but this research still focuses on mitigation data, not on prevention. [4]. Dwitriet.all[5] used k-means to grouping COVID-19 data to find out the spread of the disease in Indonesia. Spread centroid of the results of clustering with k-means covers the areas of Jakarta, West Java, and Banten. Jakarta one of the areas that has a fairly high spread of COVID-19 data, West Java also has potential to be infected, whileBanten is quite potential. Research done by Untoroet.all[6] concludes that the number of clusters obtained from data on the spread of the COVID-19 virus in Jakarta is only two clusters, even though the Indonesian government has implemented 4 levels for the lockdown rules (PPKM).

Until now, no one has investigated the maximum percentage of the possibility that the number of people who have symptoms similar to COVID-19 will actually suffer from COVID-19. But at least the number of patients with symptoms similar to COVID-19 can be used by the government or interested parties to start taking preventive measures and avoid lock down. As we know, the existence of a lock down or conditional restrictions causes the community's economy to decline [7], especially for people who work in the factory sector, self-employed without a fixed monthly salary. Several writings have proven that the COVID-19 pandemic has attacked the economy of the people in Indonesia and has reduced people's purchasing power and income. Also coupled with the implementation of the lockdown in malls, schools, offices, factories and entertainment centers [8], which limits the community's space to move in increasing their income.

K-means is often used in medical purposes such to make data clustering of cancer data set [9] and find anomalies in data. Four most popular clustering algorithms: K-Means, PAM, Agglomerative Hierarchical and DIANA and these are evaluated on eight real cancer (four Affymetrix and four cDNA) gene data and simulated data set. Beside that, AHP itself is most common algorithm in decision support system, in AHP (Analytical Hierarchy Process) Algorithm, the knowledge of expert is needed to make find eigen vector by mean matrix comparison. But the problem in AHP is to find appropriate knowledge from the right person [10], so the power of survey should find the right person or group as source of expertise.

One of the spreads of COVID-19 is going home on religious holidays, traveling between cities using public transportation so that apart from the use of lock down levels based

on the number of Covid-19 patients in one area, it is not enough. It is necessary to know the condition of people with COVID-19 who are going and those who are being addressed. Beside using mitigation level, new way to prevent Covid-19 spread is using Prevention Covid-19 (PCP) Level than can be one reason to know the vulnerability of a destination to covid-19 [11].

2. RESEARCH METHOD

To determine the dominant symptom, it is done by providing surveys and literature studies in several places in West Bandung Regency. Survey activity will collect data of 6 symptoms in 11 area and priority of symptoms according to physician. The areas then being clustering to four cluster using simple K-mean algorithm, and then the four cluster obtained will be referred to four PCP (Precaution Covid-19 Pandemic) Level. Initial centroids in K-Mean algorithm is found by implementing percentile [11]. The training and test processes can be shown at Fig 1. The reason to combine these algorithms is to eliminate subjectivity in AHP algorithm. The training process has been observed in previous research [11]. This paper will continue the previous research [12] by adding traveler check permit using Naïve Bayes Algorithm. In [12] the Naïve Bayes algorithm is used to determine someone with four Covid-19-like symptoms to be quarantined or not. This paper will advance to make permit decision for traveler who has four Covid-19-like symptoms.

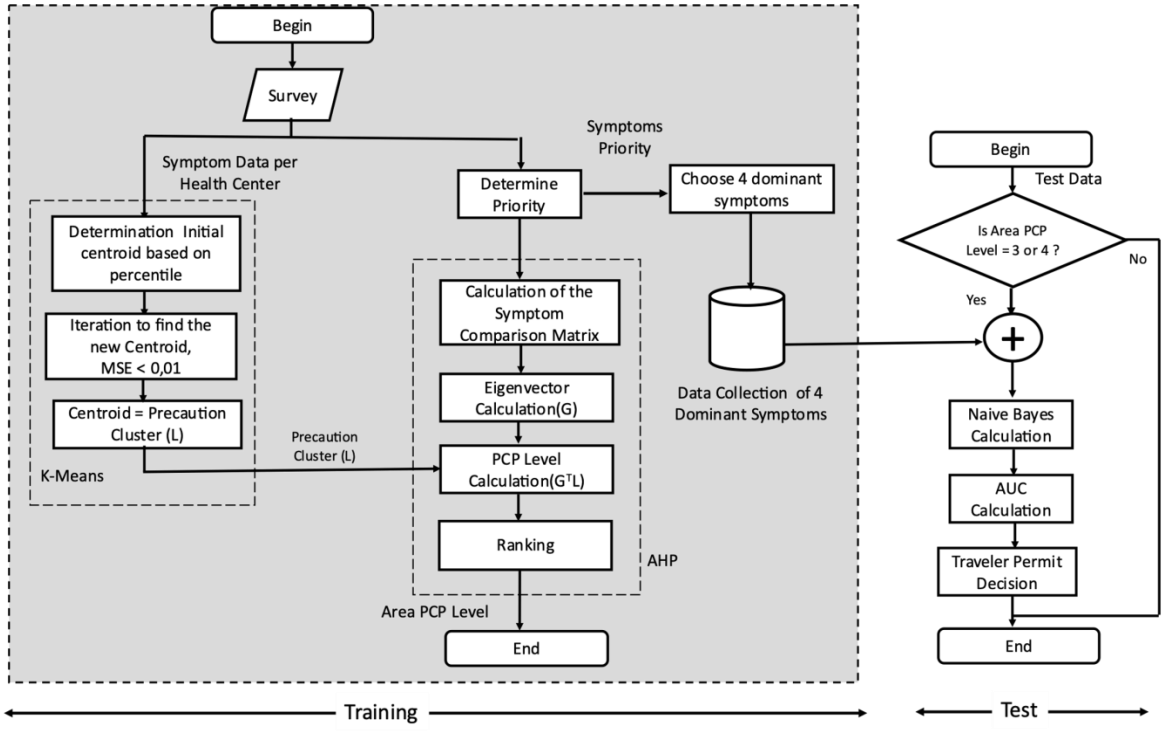

Figure 1. Research Flowchart

In determining the level of the spread of Covid-19, previously there was no guarantee information on whether the level of the Covid-19 pandemic in an area would change up or down.

Table 1. Symptoms Similar to Those of COVID-19

No. Symptoms	Name of Symptoms	Priorities
1	High fever	3
2	Dry Cough	2
3	Shoulder pain	6
4	Sore throat	1
5	Headache	5
6	Hard to breathe	4

Table 2. PCP Level

PCP Level	Description
1	No Precaution
2	Need attention
3	Need more attention
4	Intensive attention

The previous research on [11] is done with respondents are physician or doctor. But in this research the survey will be conducted to traveler. The survey process will collect the travelers information i.e. area they came from, the four dominant-symptom. The similarity of this research to previous research [11] and [12] is the usage of PCP level (Table 2) which is the prevention method beside the mitigation method that already done before new era of Covid-19Pandemic. So, this method will complete the procedure when Covid-19 pandemic has already controlled. As we know the prevention action (this research) is only used when mitigation has been controlled or overdue.

A. Calculating PCP Level

From the results of a rough survey (about 6 months during 2020) in 11 West Bandung Regencies at its Community Health Centers [12] as shown at Table 3, it was found that 6 dominant symptoms were similar to the symptoms of Covid-19 patients, namely: high fever, dry cough, shoulder pain, sore throat, headache and shortness of breath. From these symptoms, four PCP levels were determined, namely No Precaution, Need Attention, Need More Attention, Intensive Attention and severe. The severe level means that intensive strategic action is needed from the government and medical officers to prevent him from becoming a Covid-19 patients. PCP Level is calculating using K-Means algorithm with initial centroid is find from 4 percentile of average (Table 3) that is 15%, 25%, 50% and 85 percentiles.

B. Determine Traveler Permission using PCP Level

Validation is the process of evaluating the accuracy of the results of the prediction model. K-Fold Cross Validation is a validation technique that divides the data into k parts and then each part will be the classification process. By using K-Fold Cross Validation experiments will be carried out as many as k. Each experiment will use one data testing and K-1 part will become a training data, then the testing data will be exchanged with

one of training data, so for each experiment will be get a different data testing. To test the model, use the method Confusion Matrix, and the ROC curve.

1. Confusion Matrix used to analyze how well the classifier can identify tuples of different classes.

2. ROC Curve
   ROC curve (Receiver Operating Characteristic) shows the classification accuracy and compare visually. ROC express confusion matrix. ROC is a two-dimensional graph with false positives as horizontal lines and true positives to measure the performance difference method is used. ROC curves are used to measure the AUC (Area Under the Curve). AUC was calculated to measure the difference performances method was used. ROC curve divides the positive results in the y-axis and the negative results in the x-axis [13]. So the larger the area under the curve, the better the prediction results. AUC values were divided into several groups [12] modified from [14] :
   - 0.90 -1.00 = FREE TO TRAVEL WITHOUT PRECAUTION
   - 0.80 -0.90 = FREE TO TRAVEL WITH PRECAUTION
   - 0.70 -0.80 = MAY TRAVEL BETTER NOT IF NOT URGENT
   - 0.60 -0.70 = LESS PROHIBITON TO TRAVEL
   - 0.50 -0.60 = PROHIBITON TO TRAVEL

3. RESEARCH RESULTS AND DISCUSSION

From the results of the survey, the order of dominance starting from the highest to the lowest for dominant symptoms similar to those of Covid-19 patients is as follows: sore throat, dry cough, high fever, shortness of breath, headache, and shoulder pain. Of course, the order of dominance of symptoms that are similar to the symptoms of Covid-19 patients only passes at the research site being observed, while for other areas it is necessary to conduct another survey. The order of dominance of these symptoms will later be used to determine the AHP matrix, namely the comparison matrix between symptoms (Table 4) . So that the matrix is actually obtained from the survey results and reduces or even eliminates subjectivity.

Table 3. Number of people who have symptoms in 50 people

Community Health Centers	Symptoms						Avg	Init. Cent-roid
	1	2	3	4	5	6		
1. JayaMekar	4	3	2	4	1	3	2.83	
2. Tagogapu	1	2	3	5	5	5	3.50	
3. Padalarang	2	4	2	1	3	4	2.67	
4. Ngamprah	4	1	5	4	3	4	3.50	4
5. Cimareme	2	1	5	2	3	4	2.83	3
6. Padasuka	1	2	1	1	3	1	1.50	1
7. Rajamandala	2	2	1	1	1	2	1.50	

Community Health Centers	Symptoms						Avg	Init. Cent-roid
	1	2	3	4	5	6		
8. Cipendeuy	5	2	5	2	1	3	3.00	
9. Cipatat	5	3	5	2	4	3	3.67	
10. Batujajar	4	5	1	3	1	2	2.67	2
11. Cihampelas	3	5	4	3	2	1	3.00	

The calculation of the eigenvalues can be done with the arithmetic mean if the values are in almost the same order, but it can also be done using the geometric mean. Although this study does not show the comparison of the eigenvalues of the arithmetic mean (as shown at Table 5) with the geometric mean. However both have been done in this study.

As shown on Table 2 and Table 6, the area should be considered is PCP level 3 and 4. Precaution should be done to seven area in level 3 and 4. They need special attention as the biggest candidates to become Covid-19 patients, so the government must start more that seven area to prevent become Covid-19 patients

Table 4. Comparison Matrix for AHP Algorithm

	Symptom 1	Symptom 2	Symptom 3	Symptom 4	Symptom 5	Symptom 6
Symptom 1	1.000	0.500	3.000	0.333	2.000	2.000
Symptom 2	2.000	1.000	4.000	0.500	3.000	3.000
Symptom 3	0.333	0.250	1.000	0.200	0.500	0.500
Symptom 4	3.000	2.000	5.000	1.000	4.000	1.000
Symptom 5	0.500	0.333	2.000	0.250	1.000	0.500
Symptom 6	0.500	0.333	2.000	0.250	2.000	1.000

Table 5. Eigen Value and Eigen Vector using arithmetic mean

	Symptom 1	Symptom 2	Symptom 3	Symptom 4	Symptom 5	Symptom 6	Sum	Eigen Vector
Symptom 1	0.136	0.113	0.176	0.132	0.160	0.250	0.968	0.161
Symptom 2	0.273	0.226	0.235	0.197	0.240	0.375	1.547	0.258
Symptom 3	0.045	0.057	0.059	0.079	0.040	0.063	0.342	0.057
Symptom 4	0.409	0.453	0.294	0.395	0.320	0.125	1.996	0.333
Symptom 5	0.068	0.075	0.118	0.099	0.080	0.063	0.502	0.084
Symptom 6	0.068	0.075	0.118	0.099	0.160	0.125	0.645	0.107

Table 6. Training Data of Traveler related to travel permission

Case	Symptom 1	Symptom 2	Symptom 3	Symptom 4	Permit to travel
1	not often	Light	Light	Moderate	No
2	not often	Light	Light	Intermediate	No
3	not often	Light	Severe	Moderate	No
4	not often	Severe	Light	Moderate	No
5	not often	Severe	Severe	Moderate	Yes
6	not often	More Severe	Light	Moderate	Yes
7	not often	More Severe	Severe	Moderate	Yes
8	not often	More Severe	Severe	Intermediate	Yes
9	often	Light	Light	Moderate	Yes
10	often	Light	Light	Intermediate	Yes
11	often	Light	Severe	Intermediate	Yes
12	often	Severe	Light	Moderate	Yes
13	often	Severe	Light	Intermediate	Yes
14	often	Severe	Severe	Medium	Yes
15	often	More Severe	Light	Medium	Yes
16	often	More Severe	Severe	Intermediate	Yes

Calculating Traveler Permission

The Indonesian government sets quarantine rules based on the level of the number of people affected by the COVID-19 pandemic, but again this is a mitigation in the community that must be quarantined or locked down. This approach is earlier, namely by noting how likely it is that a dominant symptom similar to a Covid-19 patients in someone causes the patients to be quarantined as shown treatment. If viewed from the point of view of prevention, a deeper investigation is carried out, namely what symptoms exist at Table 7. So 4 dominant symptoms were selected from the survey results in people who were quarantined to find out how much they had. Each level of this symptom will be used as data to determine traveler permission using the Naive Bayes algorithm [12].

Table 7. PCP Level Result and Related Area

		Symptom 1	Symptom 2	Symptom 3	Symptom 4	Symptom 5	Symptom 6	Total	PCP Level	Area
Eigen Value		0.161	0.258	0.057	0.333	0.084	0.107	1.000		
Cluster area of Prevention	L1	1.667	2.667	1.333	1.000	2.333	2.333	1.811	1	Padalarang, Padasuka, Rajamandala
	L2	3.667	4.333	2.333	3.333	1.333	2.000	3.277	4	Jayamekar, Batujajar, Cihampelas
	L3	2.000	1.000	5.000	2.000	3.000	4.000	2.212	2	Cimareme
	L4	3.750	2.000	4.500	3.250	3.250	3.750	3.133	3	Tagoapu, Ngamprah, Cipendeuy, Cipatat

The naive bayes test results from Table 7 will determine whether a person is prohibited or permitted to travel. The permission check process will be implemented when traveler is going to come to an area whose PCP value is 3 or 4 . In this Naïve Bayes problem, it is determined the relationship between a person's symptoms and the PCP level of the city he is going to. Hence, the data collected from travelers are area destination and four dominant Covid-19-Like symptoms. The traveler permission process will use only four dominant Covid-19-like symptoms.

Table 8. Frequency of symptoms found intraining data

	Traveller Permit to go	
	Number	Probability
yes	12	0.75
no	4	0.25

The level of dominant 4 Covid-19-like symptoms is expressed in 2 or 3 levels, for example for shortness of breath the level is ‘not frequent’ and ‘frequent’, while for dry cough symptoms the levels are ‘mild’, ‘severe’ and ‘very severe’ as shown in the Table 7. The testing process needs to be corrected when the probability is zero, namely using the Laplacian correction as in the example in the Table 9.

Table 9. Sample Test

a. Testing Process

Example Test Data	Symp.1 Often	Symp.2 More Severe	Symp.3 Light	Symp.4 Inter-mediate
Without Correction				
Probability Yes	0.000	0.417	0.500	0.417
Probability No	0.000	0.000	0.750	0.250
With Correction (Laplacian)				
Probability No	0.250	0.250	0.750	0.625

b. Decision Process

Condition		Decision
Total of Probability for ‘no’	0.000	2
Total of Probability for ‘yes’	0.029	1

Table 10. Confusion Matrix for the Naïve Bayes Model of Covid Transmission

	True Positive	True Negative	Precision
Positive Prediction	90	21	81%
Negative Prediction	10	79	89%
Class Recall	90%	79%	

Table 11. Validation of the Naïve Bayes Model

Successful Positive Classification	Successful negative classification	Model Accuration	AUC
90%	79%	85%	0.878

Using Table 10 and calculating the accuracy and AUC values (Table 11), it can be seen that the test using the Naïve Bayes classification model in this study resulted in a good classification, with an accuracy value of 85%, and an AUC value of 0.878 which included FREE TO TRAVEL WITHOUT PRECAUTION

4. CONCLUSION

From the data provided, a method for determining candidate Covid-19 symptoms obtained for the sampling area of West Bandung Regency using the Naïve Bayes algorithm was obtained. The model for determining Covid-19 transmission based on four candidate symptoms has 89% precision and 85% accuracy. Although the accuracy and precision have not reached the value of 90% but based on the classification of Gorunescu[14] it is included in the Good Classification.

5. REFERENCES

[1] InstruksiMenteriDalamNegeriTentangPemberlakuanPembatasanKegiatanMasyarakat," Indonesian Ministry of Home Affairs, Jakarta, 2021.

[2] T. D. Baskan, "Analyzing The Going Concern Uncertainty During The Period of Covid-19 Pandemic in Terms of Independent Auditor's Report," ISPEC International Journal of Social Sciences & Humanities, vol. 4, no. 2, pp. 28-42, 2020.

[3] Hikmawati and R. Setiyabudi, "Epidemiology of COVID-19 in Indonesia: common source and propagated source as a cause for outbreaks," Journal of Infection in Developing Countries, vol. 15, no. 5, pp. 646-652, 2021.

[4] D. Abdullah, S. Susilo, A. Saleh Ahmar, R. Rusli and R. Hidayat, "The application of K means clustering for province clustering in Indonesia of the risk of the COVID 19 pandemic based on COVID 19 data," Quality & Quantity, vol. 56, no. 3, pp. 1283-1291, 2022.

[5] N. Dwitri, J. A. Tampubolon, S. Prayoga, F. I. Zer and D. Hartama, JurnalTeknologiInformasi, vol. 4, no. 1, pp. 128-132, 2020.

[6] M. CahyoUntoro, L. Anggraini, M. Andini , H. Retnosari and M. A. Nasrulloh, "Implementation of k-means method for COVID-19 data clustering in Jakarta Province," JurnalIlmiahSistemInformasi, vol. 11, no. 2, pp. 58-88, 2021.

[7] M. TarunaHabir and W. Wardana, "COVID-19's Impact on Indonesia's Economy and Financial Markets," ISEAS YusofIshakInstutute, Singapore, 2020.

[8] Update on PPKM Policy in Indonesia Effective as of March 22, 2022," Seek Legal Consultant, Jakarta, 2022.

[9] B. Hossen and R. Auwul, "Comparative Study of K-Means, Partitioning Around Medoids, Agglomerative Hierarchical, and DIANA Clustering Algorithms by Using Cancer Datasets," Biomedical Statistics and Informatics, vol. 5, no. 1, pp. 20-25, 2020.

[10] I. D. Sensue, E. Cahyaningsih and W. C. Wibowo, "Identifying Knowledge Management Process of Indonesian Government Human Capital Management Using Analytical Hierarchy Process and Pearson Correlation Analysis," in The Third Information Systems International Conference, Jakarta: ISICO2015, 2015.

[11] Y. Handoko Putra, N. Abdul Rahman, Z. Ahmad Zukarnain and R. Wahdiniwaty, "Determination of the Precaution Covid-19 Pandemic (PCP) Level Based on Dominant and Priority of Symptoms," in INCITEST 2022, Bandung, 2022.

[12] Y. Handoko Putra, N. Abdul Rahman, Z. Ahmad Zukarnain and R. Wahdiniwaty, "PenentuanPenularan Covid-19 berdasarkanGejalaKandidat Covid-19 MemanfaatkanAlgoritma Naïve Bayes," in Seminar Nasional SainsdanTeknologi ke-2, Semarang, 2022.

[13] Witten, Ian H; Frank, Eibe; Mark, A; Pal, Christopher J;, J. Data Mining: Practical Machine Learning Tools and Techniques, Fourth Edition., Cambridge, USA: Morgan Kaufman, 2017.

[14] Gorunescu, Florin, Data Mining: Concepts, Models and Techniques, Springer-Verlag Berlin Heidelberg, 2011.