Multicriteria Recommender System: A New Model for Determining risk level of COVID-19 in Indonesia

Erna Hikmawati (erna@pasim.ac.id)
Bandung Institute of Technology: Institut Teknologi Bandung
Nur Ulfa Maulidevi
Bandung Institute of Technology: Institut Teknologi Bandung
Kridanto Surendro
Bandung Institute of Technology: Institut Teknologi Bandung

Research

Keywords: Recommender System, Multicriteria, lockdown, risk, COVID-19

Posted Date: January 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-152475/v1

Multicriteria recommender system: A New Model for Determining risk level of covid-19 in Indonesia

Erna Hikmawati[1]*, Nur Ulfa Maulidevi[2], Kridanto Surendro[3]

School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Indonesia

**Corresponding author: erna@pasim.ac.id*

Abstract

Coronavirus has spread all over the world and brought significant impact to everyday life. This highly infectious virus calls for new strategies and policies that are able to suppress its transmission rate. One of the ways is by lockdown. However, lockdown is not an infallible solution and it brings with it negative impacts on society in terms of socioeconomic and mental health aspects. In addition, several factors should also be taken into account before the implementation such as demographic conditions, health services and COVID-19 case data itself. In general, a lockdown decision can be determined by looking at the level of risk of an area. Naturally, this assessment of levels of risk must be based on parameters affecting the lockdown. In this paper, we propose a multicriteria recommender system model that can calculate the value and level of risk for each region by taking into account several parameters from different databases. This model is also equipped with the process of weighting using the Analytical Network Process (ANP) method to determine interdependence and feedback between parameters. From the experiment conducted on 27 cities and districts in West Java, it was found that 15% of the regions were in the high-risk category, 41% of the regions have medium risk and 44% of the regions have a low risk.

Keywords: Recommender System, Multicriteria, lockdown, risk, COVID-19

1. Introduction

The coronavirus which has been the source of trepidation for a while now was first observed from Wuhan in December 2019 [1,2]. Coronavirus can easily spread through droplets from nose or mouth of an infected person. The droplets may land on other objects that are inadvertently contact inadvertently with other persons. The person then out of reflex touched their nose or mouth thereby contracting the virus [3]. In March 11, 2020, the virus was declared as a global pandemic by WHO [4–6]. Until Januari 10, 2021, it has spread to 223 countries with 88,828,387 confirmed cases and 1,926,625 deaths [6]. While in Indonesia itself until January 11, 2021, there were 836,718 positive cases with 24,343 deaths [7]. Never before has the scale of losses been on this level and according

to research, it is going to take more than a decade for the world to recover socially and economically [4]. The impacts are not limited to health aspects but also to other life aspects which require immediate handling to suppress its transmission.

As long as a vaccine has not been invented, every nation only has 2 options of circumventions, namely by implementing social distancing and strengthen the immune system of its citizens [8]. In addition, there is another policy that is considered the most effective in reducing the spread of the corona virus, namely by implementing lockdowns [9–13]. Lockdowns are proven to be an effective strategy in slowing down the propagation of the SARS-CoV-2 disease (infection and mortality rates) exponentially [12]. There are other studies that suggested that lockdowns will be effective in locations where the percentage of symptomatic infections is higher than the population [11]. In addition, there have been observations on lockdowns showing that the spread of the virus can be significantly reduced [12].

While lockdowns are one effective way to suppress the transmission rate of coronavirus, it also presents many negative impacts. Lockdowns implemented in the most parts of the world come with high economic and social costs [14]. In addition, the surge in worsened mental health cases have also been observed such as stress, anxiety, depression symptoms, insomnia, anger and fear in global scale [14]. A study showed that society that are subjected to a prolonged period of lockdown exhibit a lower sense of cooperation and a higher sense of egoism [15]. This is due to lessened frequency and intensity of social interactions during a lockdown. Lockdowns by the government will affect psychology, environment and economics, in addition to COVID-19 itself [13]. Therefore, lockdowns should be implemented at different levels and periods in accordance with the condition of an area [16].

There are many factors affecting the implementation of a lockdown, not just the number of positive cases. These factors include population demographics, population density, weather parameters, economy and health care system infrastructure [8,13,15,17,18].

From a study, vulnerability of an area that is important when preparing for mitigating the consequences of the coronavirus was defined [18]. The vulnerability was divided into five domains, socio-economy, demography, housing and hygiene, epidemiology, and health systems. The nature of how the coronavirus spreads, the rate of transmission and death due to infection depends on the demographic composition of the population, therefore, demographics should be part of the vulnerability index. There are three indicators to represent the demographic composition of a population in the context of COVID-19: proportion of population aged 60 years and over, proportion of population living in urban areas, and population density. In addition, the management of a pandemic

and its treatment capability depend on the ease of access to health facilities which should be part of lockdown determination parameters. Health sector indicators that can be used are the proportion of households with health insurance, the proportion of households that report no nearby public health facilities, and the number of public hospitals per 100,000 population (for the district level) and the number of hospital beds per 1000 population (for state or union territory level) [18].

These parameters or indicators that can be referred to to take this lockdown policy have different data sources. A method for collecting and analyzing data so that it can be used as a basis for decision making is needed. Many companies dedicate a significant chunk of their resources to collect and analyze transactional data. The analysis may come from any type of application [19]. Multi-database mining is thus an important issue in data mining research, where effective local pattern analysis and efficient mining approaches are required [19–21].

Tackling this pandemic will rely heavily on broad understanding and mass participation by the entire global community, not just health professionals and high-level decision makers that design responses. Currently, half of the world's population has implemented a lockdown, which means that the coronavirus has spread and must be overcome. Faced with such a very difficult decision, it is important for policymakers, health professionals and the general public that as many people as possible understand why implementing strict lockdowns is the only realistic way for individual countries to tackle their national-level epidemic before it turns into a public health disaster [22].

To support this necessity, a tool for planning and prioritizing at the district level as well as an effective allocation of resources are needed [18]. For this purpose, understanding of the transmission risks in every urban district are vital for the government and administration to implement a reopening strategy [17]. The tool in question should be able to predict the transmission risk in districts/cities to support decision makers in identifying the best strategy to reduce or restart local activities during lockdown conditions taking into account certain parameters.

In general, a lockdown decision can be taken by looking the risk level of an area. Naturally, this assessment of levels of risk must be based on parameters affecting the lockdown. Due to the numbers of parameters that influence the decision making for lockdown and also for risk assessment, this issue falls into the category of multicriteria decision problem. Multicriteria decision making is used where the selection of elements is based on several factors that help in decision making. One method that can be used in completing Multicriteria Decision Making (MCDM) is the Analytical Network Process (ANP) method [23,24]. ANP was introduced to solve interdependence and feedback between elements in a cluster of between clusters. ANP is suitable in situations where parameters are interdependent and need feedback as well [25].

There is one system that is popular in providing recommendations to someone, namely the recommender system [26–31]. It can be used to build recommendations for areas where lockdowns are indicated by looking at the risk level taking into account several weighted parameters from many databases [19,32]. In this paper, we will propose a multi-criteria recommender system model that is derived from many databases and involves various criteria that have been weighted using the ANP method which is an implementation of the model previously presented [33]. In addition, this model is also a small part of the adaptive rule model in terms of establishing the recommender system that the author has previously suggested [34]. Related works will be presented in part II, proposed methods will be described in section III, experimental results and discussion will be explained in section IV, and conclusions will be explained in section V.

2. Related Work

2.1. Recommender System

Currently, recommender system has been widely used in various fields. The development of this method has been mostly carried out based on its basic methods, namely collaborative filtering, content-based filtering and hybrid recommender systems.

There is a recommender system method that is developed based on a user preference model by looking at explicit and implicit characteristics according to user interests. This method is known as the Collaborative Filtering Method. Explicit indicators are usually user ratings of an item, while implicit indicators can be the length of time users have spent interacting with the content or the level of interaction [27].

In addition, there is a recommender system that works by identifying similarities between items based on their descriptions. Items that have a similar description will be used as recommendations for target users. This method is known as the Content Based Filtering Method [27].

In addition to methods abovementioned, currently there is a development of recommender systems that use association rule mining [27–30,35]. By following the development of the recommender system method, there is still an open possibility for the development of a recommender system that integrate other methods to produce more effective recommendation results.

2.2. Multicriteria Analysis

Multicriteria Analysis is a decision-making method used in which the selection of elements is based on several factors that help in decision making [23,24]. Several methods that can be used in multicriteria analysis are the Analytical Hierarchy Process (AHP) and the Analytical Network Process (ANP). ANP is a development of AHP. The difference between AHP and ANP lies in the structure of the model, where AHP uses a hierarchical structure while ANP uses a network structure. This is what makes ANP able to handle interactions and feedback relationships between criteria/sub-criteria and alternatives [23–25,36].

In addition, the multicriteria analysis method may also use Electree method [37,38]. The electree method can rank the set from best to worst. It is a multi-criteria decision-making method based on the concept of outranking using pairwise comparisons of alternatives based on each appropriate criterion [37,38]. The concept of outranking is an alternative that is dominated or dominates other alternatives, so there is no structure in this method.

2.3. COVID-19 Risk Parameters

In determining the risk level of COVID, many parameters need to be taken into account. Several previous studies have revealed the parameters associated with this risk of COVID. In those papers [17], parameters involved in the risk of COVID transmission were described and classified into criteria, sub-criteria and intensity as shown in Figure 1.

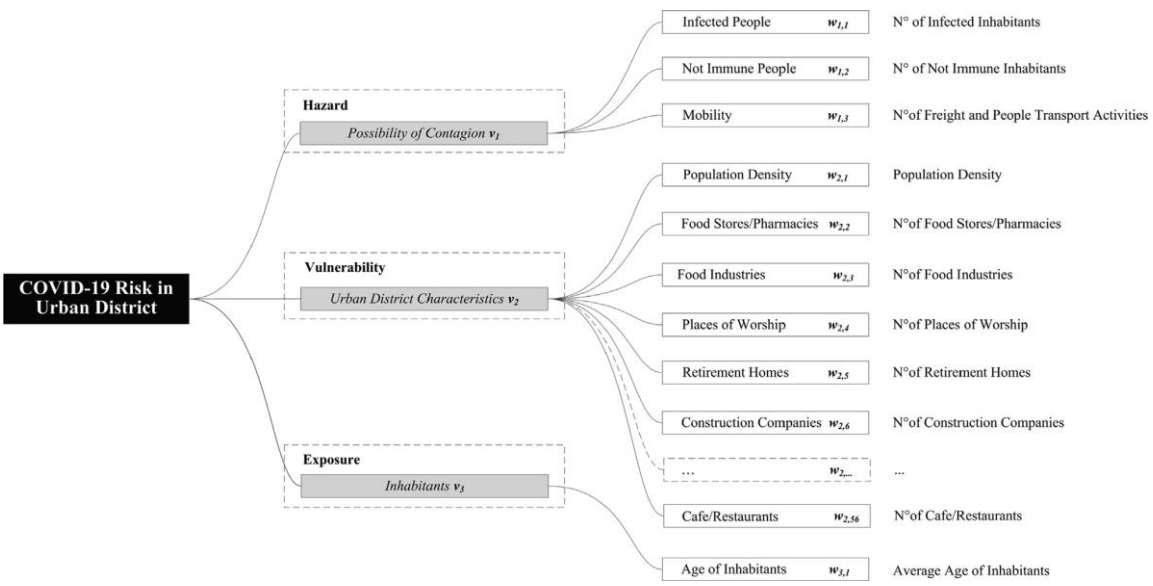

Fig.1 Structure of the problem in Criteria, Sub-Criteria and Intensity [17]

In figure 1, there are criteria divided into 3 categories, namely Hazard, Vulnerability and Exposure with the following explanation:

1. Hazard takes into account the possibility of transmission and it is characterized by three sub-criteria namely infected person, non-immune person and mobility.

2. Vulnerability evaluates the characteristics urban districts by taking into account fifty-six sub-criteria, including population density and the existence of crowded places such as pharmacy, food industry, places of worship, retirement homes, construction companies, open cafes/restaurants, etc.

3. Exposure evaluates resident typology who can suffer fatal consequences by considering the age of population.

In addition, Acharya [18] in his paper also proposed the vulnerability index that affects COVID. Acharya calculated a composite vulnerability index at the state and district level based on 16 indicators categorized into the following five domains: socioeconomic, demographic, housing and hygiene, epidemiology, and health system. Afterward, the percentile rank method is used to calculate the domain-specific and overall vulnerabilities as well as present the results spatially with the number of COVID-19 positive cases in sub-districts. The vulnerability indicators presented can be seen in figure 2.

	Variable description	Source
Socioeconomic		
Scheduled tribe or caste households	Calculated as proportion of households belonging to scheduled caste or tribe	National Family Health Survey-4 (Household file), India (2015–16)[6]
Education level in population	Calculated as proportion of population who completed secondary or higher level of education	National Family Health Survey-4 (Person file), India (2015–16)[6]
Poor households	An asset deprivation indicator was computed as the proportion of households that did not have any of the following: a motorised vehicle (a two-wheeler, car or truck, or tractor), television, computer, bicycle, refrigerator, thresher, or air-conditioner or cooler	National Family Health Survey-4 (Household file), India (2015–16)[6]
Demographic		
Elderly population	Calculated as proportion of individuals in the population aged 60 years or older	National Family Health Survey-4 (Person file), India (2015–16)[6]
Urbanisation	Calculated as proportion of urban households among all households	National Family Health Survey-4 (Household file), India (2015–16)[6]
Population density	Calculated as a ratio of population of a unit (district or state) and its area in km^2	Area data: 2011 census;[19] population data: linearly projected population for 2019 using growth rate calculated for each district based on 2001 and 2011 census
Housing and hygiene condition		
People per room	Calculated as the mean number of people residing per room used for sleeping in a household	National Family Health Survey-4 (Household file), India (2015–16)[6]
Households with no toilet facility	Calculated as proportion of households reporting no availability of toilet facility within premises	National Family Health Survey-4 (Household file), India (2015–16)[6]
Households with no hand-hygiene facility	Calculated as percent households with no availability of water and soap or detergent at place of handwashing	National Family Health Survey-4 (Household file), India (2015–16)[6]
Availability of health care		
Households with health insurance	Calculated as proportion of households with at least one member covered under any health insurance scheme	National Family Health Survey-4 (Household file), India (2015–16)[6]
Households without easy access to public health facility	Calculated as proportion of households reported having no nearby public health facility	National Family Health Survey-4 (Household file), India (2015–16)[6]
Availability of public hospitals (at district level)	Calculated as number of public hospitals (primary health centre and above) per 100 000 population	Rural health statistics 2018 and linearly projected population for 2019 using growth rate calculated for each district based on 2001 and 2011 census[20]
Availability of hospital beds (at state level)	Calculated as number of public or private hospital beds per 1000 population	National health profile 2019[21]
Epidemiological		
Men with any chronic morbidity	Calculated as proportion of men aged 40–54 years with chronic health conditions, such as cardiovascular disease, diabetes, asthma, or cancer	National Family Health Survey-4 (Men's file), 2015–16[6]
Men who smoke	Calculated as proportion of men who smoke tobacco	National Family Health Survey-4 (Men's file), 2015–16[6]
Women with any chronic morbidity	Calculated as proportion of women aged 40–49 years with chronic health conditions, such as cardiovascular disease, diabetes, asthma, or cancer	National Family Health Survey-4 (Women's file), 2015–16[6]

Fig.1 Domains of vulnerability and variables within [18]

In Indonesia itself, a regional risk zoning map has been presented which is calculated based on public health indicators using scoring and weighting. The indicators used are as follows [7]:

1. EPIDEMIOLOGICAL INDICATORS:
    a. ≥50% decrease in the number of positive & probable cases in the last week from the peak
    b. ≥50% decrease in the number of suspected cases in the last week from the peak
    c. ≥50% decrease in the number of deaths from positive & probable cases in the last week from the peak
    d. ≥50% decrease in the number of deaths in suspected cases in the last week from the peak
    e. ≥50% decrease in the number of positive & probable cases hospitalized in the last week from the peak
    f. ≥50% decrease in the number of suspected cases hospitalized in the last week from the peak
    g. The cumulative percentage of recovery cases from all positive & probable cases
    h. Incidence rate of positive cases per 100,000 population
    i. Mortality rate of positive cases per 100,000 population
    j. Incidence rate of positive cases per 100,000 population
2. COMMUNITY HEALTH SURVEILLANCE INDICATORS
    a. The number of diagnostic samples examined increased over the past 2 weeks
    b. Low positivity rate (target ≤5% positive sample of all people examined)
3. HEALTH SERVICE INDICATORS
    a. The number of beds in the Referral Hospital isolation room can accommodate up to >20% of the number of COVID-19 positive patients being treated in the hospital
    b. The number of beds in the Referral Hospital can accommodate up to> 20% of the number of ODP, PDP, and COVID-19 positive patients hospitalized

The indicators were derived from positive case data and laboratory examinations based on the surveillance data of the Ministry of Health, ODP and PDP patient data, and hospital service capacity is obtained based on online hospital data under the coordination of the Director General of Health Services, Ministry of Health. The visualization of the risk zone map drawn up by the Indonesian government can be seen in figure 3.

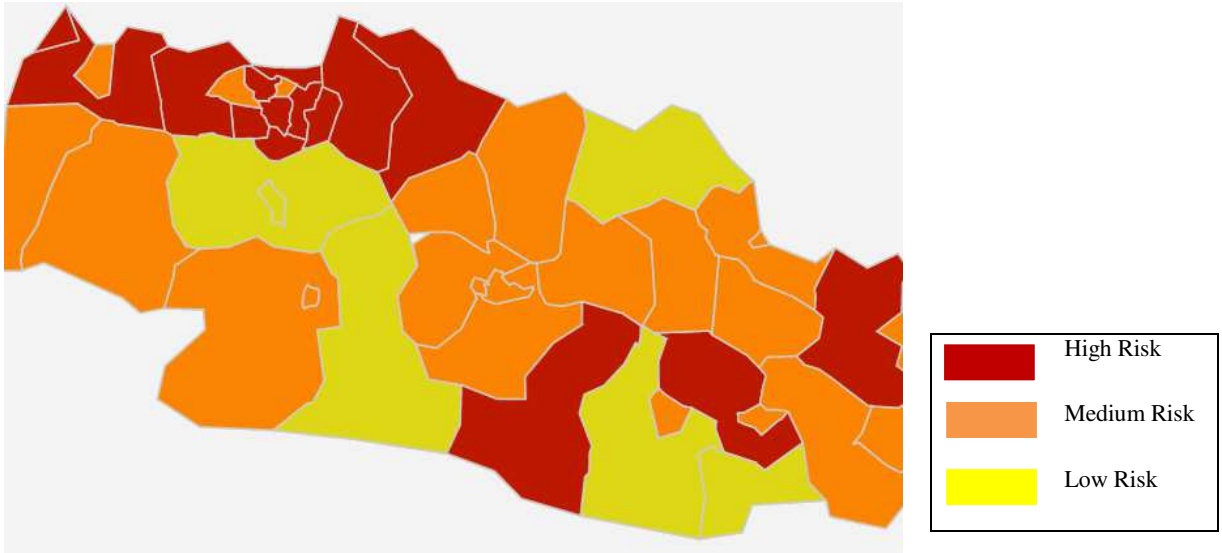

Fig.3 Covid-19 risk map for the West Java region [7]

2.4. SIR Model

One model that can be used to predict COVID-19 cases is the SIR model. The majority of the COVID-19 epidemic models come from the SIR model and many variations of it have been used in several countries, such as India, China, Italy and Brazil. The basic formula for the SIR model is [39]:

$$\frac{dS(t)}{dt} = -\beta S(t)\frac{I(t)}{N} \quad [1]$$

$$\frac{dI(t)}{dt} = \beta S(t)\frac{I(t)}{N} - \gamma I(t) \quad [2]$$

$$\frac{dR(t)}{dt} = \gamma I(t) \quad [3]$$

$$S(t) + I(t) + R(t) = N \quad [4]$$

Where:

S(t) : number of susceptible people

I(t) : number of people infected

R(t) : number of people removed

β : infection rate

ϒ : removed rate

The COVID virus can be assessed from the basic reproduction number (R0). If R0>1 then the virus is still spreading and if R0<1 then the outbreak is going to case and the number of new cases will decline. R0 can be calculated using the following formula [39]:

$$\frac{dI(t)}{dt} > 0 \Rightarrow \beta S(t)\frac{I(t)}{N} - \gamma I(t) > 0 \Rightarrow I(t)\left(\beta\frac{S(t)}{N} - \gamma\right)$$
$$> 0 \overset{S(t)\approx N}{\Rightarrow} I(t)(\beta - \gamma) \Rightarrow \beta - \gamma > 0 \Rightarrow \beta > \gamma \Rightarrow \frac{\beta}{\gamma} \qquad [5]$$
$$> 1 \Rightarrow R_0 = \frac{\beta}{\gamma}$$

This SIR Model is especially useful for policymakers to identify the potential impact of a pandemic and prompt them to take preliminary actions to minimize the impact. Moreover, after the declaration of the state of global pandemic, more information is needed for detailed planning, such as the peak arrival of the pandemic, the number of hospital beds required at peak times, and decision making to loosen/lift lockdowns, and finally return to normal life [5].

2.5. Analytical Network Process (ANP)

The Analytical Network Process is the evolution of the Analytical Hierarchy Process (AHP) method developed by Saaty [23,24]. The ANP method is used as a multicriteria decision tool that is able to solve dependency problems for various elements and feedback between layers in a network [36]. AHP lacks this capability. ANP can handle interactions and feedback relationships between criteria/sub-criteria and alternatives. ANP is suitable in situations where parameters are interdependent and need feedback as well. Furthermore, ranking criteria and alternative elements can be used for decision making [25]. The general steps for the ANP method are as follows [23,24].

1. The first step is formulating the problem, where it is identified and divided into subproblems, if needed. The objective is clearly defined as well as the criteria (parameters) and alternatives (elements). Defining the criteria/sub-criteria is important because the objective entirely rely on this parameter. Based on this parameter, alternatives are chosen.

2. The components in each cluster are compared in pairs, based on the quantitative scale proposed by Saaty as shown in table 1. Each element is scaled based on their importance on other elements by taking into account certain parameters. A matrix for each comparison is made, where 1 represents equal importance while 9 represents the most important.

Table 1 Quantitative 9-point scale [36]

Scale	Description
1	Equal relative importance
2	Equally to moderately more important
3	Moderately more important
4	Moderately to strongly important
5	Strongly important
6	Strongly to very strongly more important
7	Very strongly more important
8	Very strongly to extremely more important
9	Extremely important (high priority)

3. The results of the pairwise comparison are presented in the form of a matrix, then all the elements of each column in the matrix are added using the following formula:

   For C1 to Cn-1:

$$S_i = \sum_{j=1}^{k-1} a_{ji} + \sum_{j=k}^{n} \frac{1}{a_{ij}} \text{ where } k=2,3,4,\ldots,n \text{ and } i=1,2,3,\ldots,n\text{-}1 \quad [6]$$

   For Cn:

$$S_n = \sum_{j=1}^{k-1} a_{ji} \text{ where k=n+1, i=n} \quad [7]$$

   Afterward, matrix normalization was performed by dividing each column with the number of each column. Each row is added and the mean of each row is used to obtain Eigen Vector.

4. It is very important to check the reliability of each comparison made. To measure the consistency of Saaty's comparisons [23,24], Consistency Ratio (CR) is used to define how many of the comparisons made are consistent. CR should match or be less 0.1, which means inconsistency is allowed up to 10%. If it is over that, the comparisons need revision. Consistency Measure (CM) is the first step in conducting a consistency analysis. CM vector is the input for consistency index calculation and ratios (CI and CR). To identify CM, the matrix that has been normalized with Eigen vector is divided with element EV as shown in figure 4.

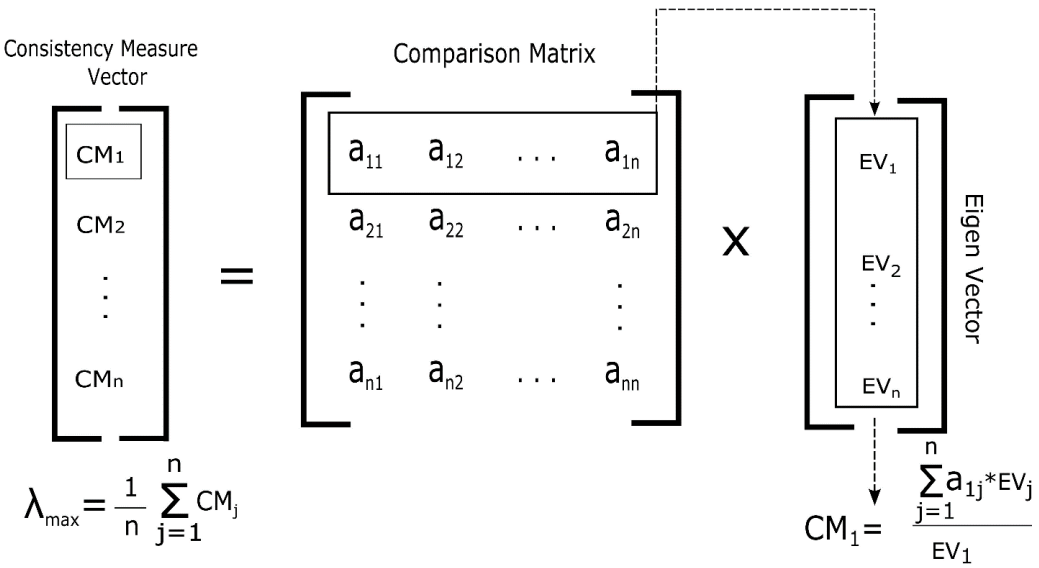

Fig.4 Consistency Measure calculation (CM) [36]

The general form to obtain CM is given in Equation (8), where Rj is the comparison rows that match the matrix, EV is Eigen vector (priority vector) and EVj represents corresponding elements in EV. The mean of the CM vector is λmax.

$$CM_j = \frac{R_j \times EV}{EV_i} \text{ where j=1,2,3,…., n} \qquad [8]$$

Furthermore, to calculate the Consistency Index, namely the degree or deviation of consistency, you can use the formula:

$$CI = \frac{\lambda_{\max} - n}{n-1} \qquad [9]$$

To check the reliability of the pairwise comparisons, it is highly important to ensure the consistency between pairwise comparisons made. To obtain CR, we need to calculate consistency index (CI) as shown in Equation (9) and RI (RI values are drawn from Table 2 according to the order of the matrix represented by n).

Table 2 Random Index (RI)

N	1	2	3	4	5	6	7	8	9
RI	0	0	0.52	0.89	1.11	1.25	1.35	1.40	1.45

To calculate the CR value, a formula can be used:

$$CR = \frac{CI}{RI} \qquad [10]$$

5. The results of all comparison matrices are merged into an unweighted supermatrix. This local priority is converted into a weighted supermatrix by making it a stochastic column.

6. The weighted supermatrix is turned into a matrix boundary by promoting it to 2k power in order to be a more stable value, where k is any number. The matrix boundary is a resultant matrix, containing the final weight of each element. It determines the best alternatives and the most important criteria as well.

7. A sensitivity check is performed to determine the stability of the alternative ratings.

The general steps of ANP can also be seen in figure 5.

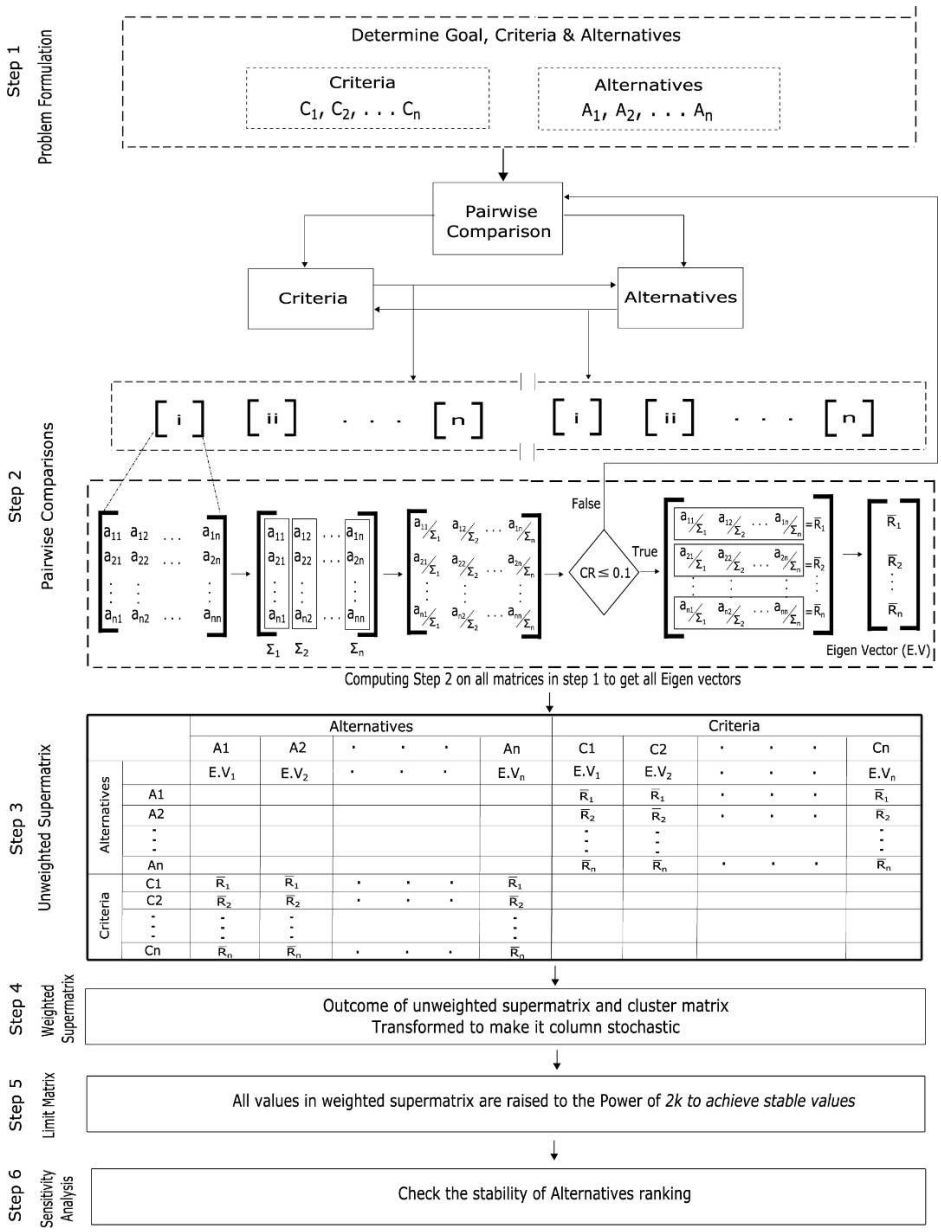

Fig.5 General Step for ANP [36]

3. The Proposed Methods

3.1. Data

The case study in this study was COVID-19 cases in West Java at the city/district level. In this study there were 7 parameters used in calculating the risk level of COVID in an area. The parameters were derived from various databases. The list of parameters and their sources can be seen in table 3.

Table 3 Parameter dan Sumber Data

Parameter	Source
Reproduction Number (R)	West Java COVID-19 case statistics data from the Covid-19 Information and Coordination Center of West Java Province [40]
Number of Health Facilities (FK)	West Java Health Profile Data from the West Java Health Office [41]
Number of residents who are elderly (LU)	West Java Health Profile Data from the West Java Health Office [41]
Number of Beds in Health Facilities (TT)	West Java Health Profile Data from the West Java Health Office [41]
Number of Nurses (P)	West Java Health Profile Data from the West Java Health Office [41]
The number of medical personnel consisting of general practitioners and specialist doctors (TM)	West Java Health Profile Data from the West Java Health Office [41]
Population density (KP)	Population Data from the Central Bureau of Statistics [42]

3.2. Method

The Multicriteria Recommender System model used in this study was an implementation and evolution of the previous recommender system model [33]. It was also a small part of the adaptive rule model [34] that the author had compiled previously. The development on the multicriteria recommender system model was performed by adding the ANP method in the weighting. The proposed model can be seen in figure 6.

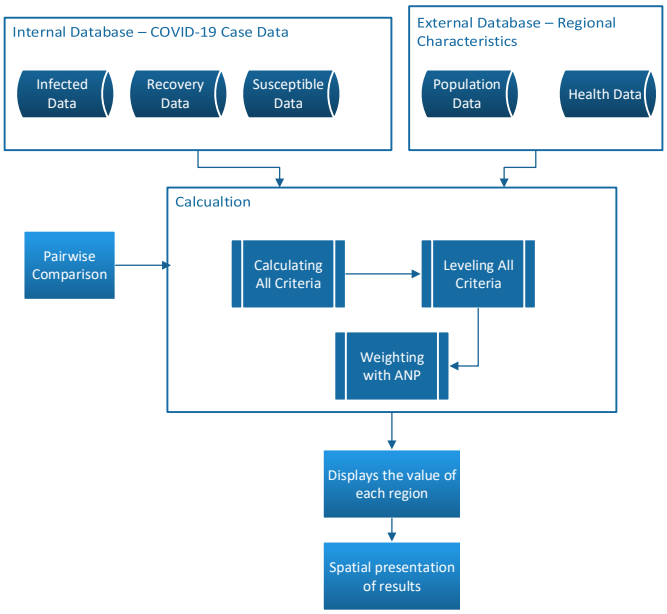

Fig.6 Model Multicriteria Recommender System

In the model, there were two input categories, namely internal database and external database. The internal database is a database that was directly related to the topic of recommender systems, in this case was COVID case data taken from the center of information and coordination of COVID-19 in West Java. The external database was a supporting database that influenced the calculation of the risk of COVID in an area. Health data and population density data were the external databases in this study. The two input categories that had been previously described were used as criteria in the weighting process using the ANP method. Therefore, pairwise comparisons for each of the criteria referred to were needed.

The main process in this multicriteria recommender system model was the process of calculating the risk level of COVID in cities and districts in West Java. There were three main sub-processes in the multicriteria recommender system model, namely:

1. The calculation of all criteria.

   At this stage, the values for all parameters for each city/district were calculated according to the formula used respectively. The R value was calculated from the derivative of the SIR formula. For the old age, bed, nurses, medical personnel parameters, ratio to 1000 inhabitants was calculated.

2. The levelling of all criteria

   After obtaining the numbers for each parameter, data categorization was carried out and divided into three categories, namely low, medium and high with the help of standard deviation values.

3. The weighting with the ANP method

   Finally, weighting was carried out for each existing criteria and alternatives according to the level calculated in the previous stage.

The main process in the multicriteria recommender system model can be seen in figure 7.

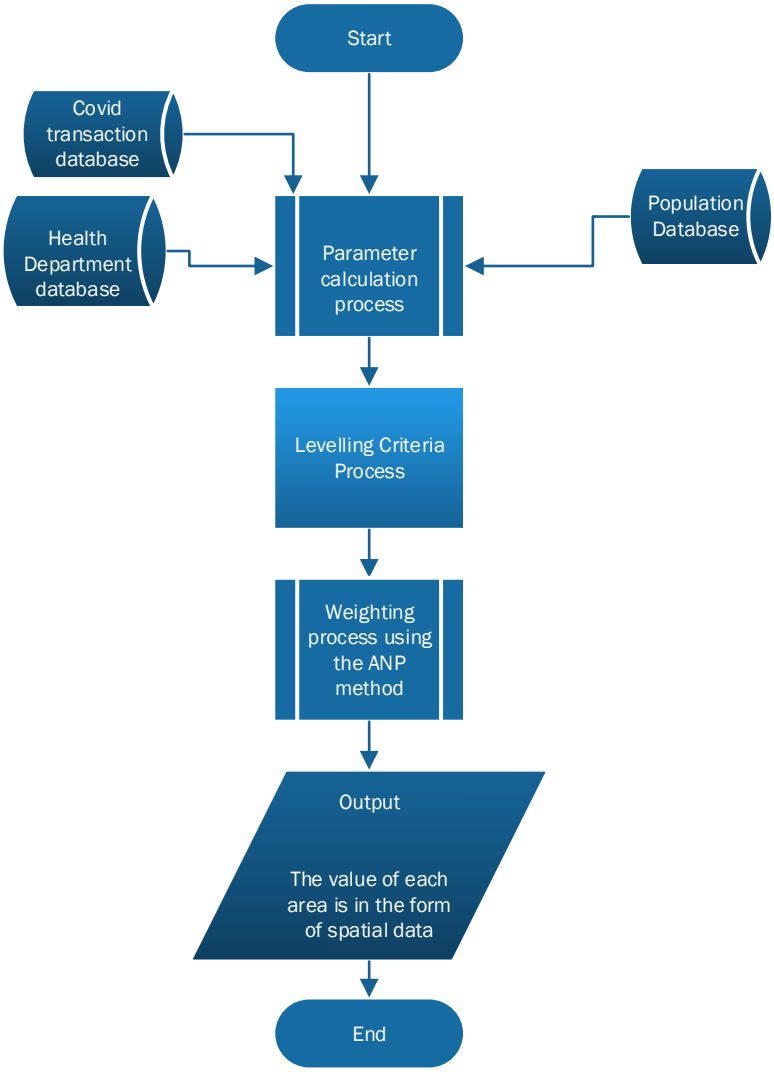

Fig.7 Main Flow of Multicriteria Recommender System Model

The end results of this model were points for each region, which were then categorized into low, medium or high level risk. To add to data visualization, each region value was presented in spatial data in the form of the map of West Java that had been sprinkled with different colors for each city/district in accordance with their respective level.

4. Result and Analysis

4.1. Preprocessing

The first stage in this study was data preparation. The steps were:

1. Calculating the R value for each city and districts using formula [5]. The COVID transaction data used was data starting from August 1, 2020 - December 23, 2020.

2. The old age, bed, nurse, medical personnel parameters were calculated against 1000 inhabitants.

3. For each parameter data categorization was performed using standard deviation. The existing data were categorized into three categories namely high, medium and low. The data categorization used the following guidelines:

Table 4. Data categorization guidelines

Category	Value
Low	X < M – 1SD
Medium	M – 1SD ≤ X < M + 1SD
High	M + 1SD ≤ X

Where:

X: data that are about to be categorized

M: Mean

SD: Standard Deviation

The standard deviation can be calculated using the formula [43,44]:

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}} \qquad [11]$$

Where:

$\bar{x}$: mean

n : the amount of data

x_i : value data

After the categorization, the initial data that would be used in this study can be seen in table 5.

Table 5. Preliminary Data

Area Code	Region Name	FK	LU	TT	P	TM	KP	R
3201	Bogor (BGR)	High	High	Low	Low	Low	High	Low
3202	Sukabumi (SKBM)	Moderate	High	Low	Low	Low	High	Low
3203	Cianjur (CJR)	High	High	Low	Low	Low	High	Low
3204	Bandung (BDG)	High	High	Low	Low	Low	High	Moderate
3205	Garut (GRT)	Moderate	High	Low	Low	Low	High	High
3206	Tasikmalaya (TSK)	Moderate	High	Low	Low	Low	High	Low
3207	Ciamis (CMS)	Low	Low	Low	Low	Low	High	Moderate

3208	Kuningan (KNG)	Low	Moderate	Moderate	Low	Low	High	Low
3209	Cirebon (CRB)	Low	High	Low	Low	Low	High	Moderate
3210	Majalengka (MJG)	Moderate	Low	Low	Low	Low	High	High
3211	Sumedang (SMDG)	Moderate	Low	Low	Low	Low	High	High
3212	Indramayu (INDR)	Low	Moderate	Low	Low	Low	High	Low
3213	Subang (SBG)	Low	High	Low	Low	Low	High	Low
3214	Purwakarta (PWKT)	Moderate	Moderate	Moderate	Low	Low	High	Moderate
3215	Karawang (KRWG)	Moderate	Moderate	Moderate	Moderate	Low	High	High
3216	Bekasi (BKS)	Moderate	Moderate	Low	Low	Low	High	Moderate
3217	Bandung Barat (BDGBRT)	High	Moderate	Low	Low	Low	High	Moderate
3218	Pangandaran (PNGNDRN)	Moderate	Moderate	Low	Low	Low	High	Low
3271	Bogor City (KBGR)	Low	High	High	Moderate	Low	Low	Moderate
3272	Sukabumi City (KSKBM)	Low	Moderate	Low	High	Low	Moderate	High
3273	Bandung City (KBDG)	Low	Moderate	Moderate	Low	Low	Low	Low
3274	Cirebon City (KCRB)	Low	High	Low	High	High	Moderate	Low
3275	Bekasi City (KBKS)	Moderate	High	High	Low	Low	Low	Low
3276	Depok City (DPK)	High	High	Moderate	Low	Low	Low	Moderate
3277	Cimahi City (CMH)	Moderate	High	High	High	Low	Low	Moderate
3278	Tasikmalaya City (KTSK)	Low	Moderate	High	Low	Low	High	Low
3279	Banjar City (BNJR)	Low	Moderate	High	Low	Low	High	Moderate

West Java regional characteristics data for each parameter can also be seen in figure 8.

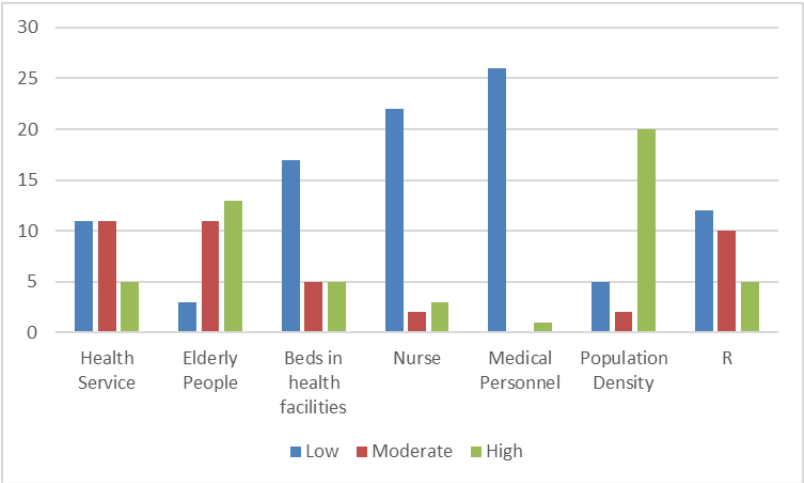

Fig.8 Regional Characteristics in West Java

4.2. ANP Process

The initial stage in the weighting process with ANP was to determine the criteria and alternatives that would be used. This research consisted of 7 criteria, namely FK, LU, TT, P, TM, KP and R, while the alternatives were 27, namely a number of cities and districts in West Java.

Set 1: C = {FK,KP,LU,P,R,TM,TT}

Set 2: A = {BDG, BDGBRT, BGR, BKS, BNJR, CJR, CMH, CMS, CRB, DPK, GRT, INDR, KBDG, KBGR, KBKS, KCRB, KNG, KRWG, KSKBM, KTSK, MJG, PNGNDRN, PWKT, SBG, SKBM, SMDG, TSK}

Afterward, pairwise comparisons were performed in accordance with the number of criteria and alternatives to ensure <1 CR value. The results of the CR value for each comparison can be seen in table 6.

Table 6. CR Value for all pairwise comparison

Node	Pairwise	CR
FK	All Alternative	0.01694
KP	All Alternative	0.00188
LU	All Alternative	0.00476
P	All Alternative	0.00124
R	All Alternative	0.00557
TM	All Alternative	0.00000
TT	All Alternative	0.00397
BDG	All Criteria	0.09375
BDGBRT	All Criteria	0.09719
BGR	All Criteria	0.09735
BKS	All Criteria	0.09467
BNJR	All Criteria	0.09067
CJR	All Criteria	0.09653
CMH	All Criteria	0.09112
CMS	All Criteria	0.09662
CRB	All Criteria	0.08816
DPK	All Criteria	0.0883
GRT	All Criteria	0.09623
INDR	All Criteria	0.09978
KBDG	All Criteria	0.09987
KBGR	All Criteria	0.07947
KBKS	All Criteria	0.08585
KCRB	All Criteria	0.08348
KNG	All Criteria	0.09106
KRWG	All Criteria	0.099
KSKBM	All Criteria	0.09948
KTSK	All Criteria	0.06739
MJG	All Criteria	0.09324
PNGNDRN	All Criteria	0.08158

PWKT	All Criteria	0.09963
SBG	All Criteria	0.09652
SKBM	All Criteria	0.09402
SMDG	All Criteria	0.08028
TSK	All Criteria	0.09589

After the pairwise comparisons, a supermatrix was made consisting of unweighted matrix, weight matrix dan limit matrix. From this limit matrix, the final value for each city and district was generated. The final results of the ANP process can be seen in table 7.

Table 7. Limit Matrix value from ANP Process

Area	Value
BDG	0.01487
BDGBRT	0.01506
BGR	0.01941
BKS	0.01608
BNJR	0.01799
CJR	0.01939
CMH	0.0148
CMS	0.02009
CRB	0.01841
DPK	0.01652
GRT	0.01383
INDR	0.2394
KBDG	0.026
KBGR	0.01808
KBKS	0.0223
KCRB	0.01761
KNG	0.02302
KRWG	0.01179
KSKBM	0.01602
KTSK	0.02294
MJG	0.01568
PNGNDRN	0.02101
PWKT	0.01533
SBG	0.02328
SKBM	0.02037
SMDG	0.01576
TSK	0.02043

If the values were categorized into 3 levels, namely low, medium and high, the overall assessment results can be seen in Figure 9.

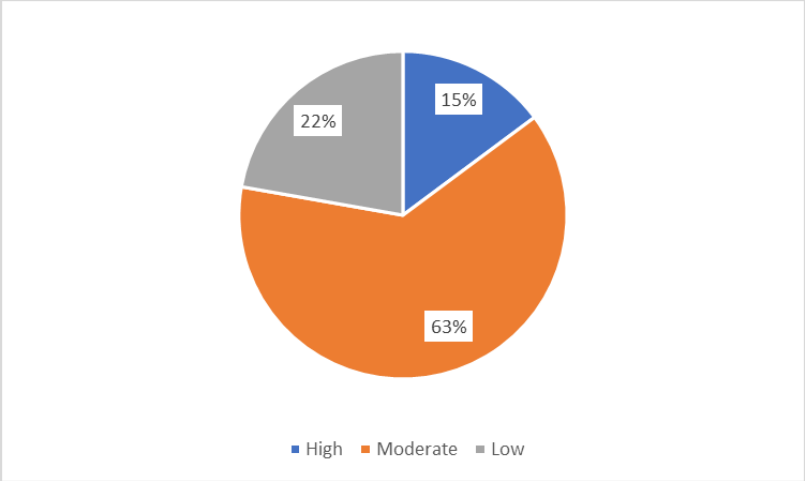

Fig.9 Percentage of area assessment results in West Java

The values generated from the ANP process would be added on the map of West Java to obtain data visualization in the form of spatial data as shown in figure 10.

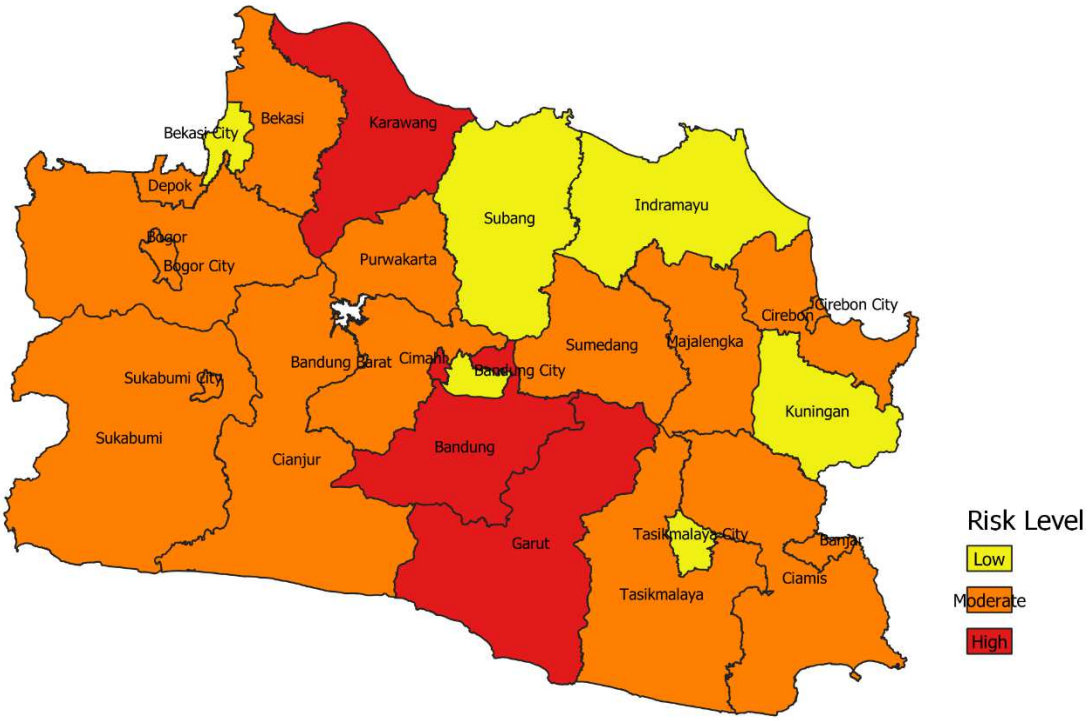

Fig.10 Spatial visualization of the final result

In figure 10, there were differences in the classification of the risk levels when compared to figure 2. This was because the indicators used were different. There were several parameters that were not used by Indonesian government namely population density, number of health facilities, number of nurses, number of medical personnel and number of elderly people. As an example, Bandung Regency was categorized as high-risk by the recommender

system because it has low lumber of nurses and medical personnel while the population density and the number of elderly people were high. Bogor City was categorized as medium-risk because it has low number of health facilities, nurses and medical personnel. However, its low population density brought the categorization to medium-risk.

5. Conclusion and Future Work

The pandemic afflicting nearly the whole world has led to significant changes in life. New rules against this pandemic must be issued immediately to suppress its transmission rate. One of the ways is by implementing lockdowns. However, lockdowns also entail negative effects on people in an area from a socio-economic and health perspective (stress, anxiety, depression symptoms, insomnia, anger and fear in a global scale). Therefore, the right strategies and policies are needed in implementing lockdowns.

A lockdown decision can be taken by looking at the level of risk of an area. Measuring the risk level of an area takes into account several factors. These factors include population demographic conditions (population density), condition of health facilities (number of health facilities, number of nurses, number of medical personnel, number of beds in health facilities, number of elderly people) and of course data on corona cases itself (reproductive numbers).

The multicriteria recommender system model is able to calculate the risk level of an area based on predetermined parameters. In order to make the calculation more effective, the ANP weighting method is incorporated to perform pairwise comparisons between criteria and alternatives. The results of the multicriteria recommender system model can be presented in the form of spatial data and used by policymakers to take decisions.

In the future, the author will attempt to develop a real-time multicriteria recommender system by integrating available data. That way, the process of updating data and results can be done automatically.

ABBREVIATIONS

ANP: Analytical Network Process

AHP: Analytical Hierarchy Process

MCDM: Multicriteria Decision Making

COVID: Corona Virus Disease

CM: Consistency Measure

CR: Consistency Ratio

DECLARATIONS

Acknowledgments

We would like to thank Institut Teknologi Bandung and Pasim National University for supporting this research.

Authors' contributions

The author confirms the sole responsibility for this manuscript fully as a sole author for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation. The author read and approved the final manuscript.

Funding

I sincerely thank LPDP (Indonesia Endowment Fund for Education), Ministry of Finance, Republic Indonesia for providing me with the financial support for this research. This research was funded by LPDP (Indonesia Endowment Fund for Education), Ministry of Finance, Republic Indonesia.

Availability of data and materials

The original data used for this study is available in:

1. West Java COVID-19 case statistics data from the Covid-19 Information and Coordination Center of West Java Province (https://pikobar.jabarprov.go.id/)
2. West Java Health Profile Data from the West Java Health Office (http://www.diskes.jabarprov.go.id/)
3. Population Data from the Central Bureau of Statistics (https://www.bps.go.id/)

Competing interests

The author reports no potential conflict of interest.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Aldila D, Khoshnaw SHA, Safitri E, Anwar YR, Bakry ARQ, Samiadji BM, et al. A mathematical study on the spread of COVID-19 considering social distancing and rapid assessment: The case of Jakarta, Indonesia. Chaos, Solitons & Fractals. 2020;139:110042.

2. Shorten C, Khoshgoftaar TM, Furht B. Deep Learning applications for COVID-19. J Big Data. 2021;8:18.

3. Annas S, Isbar Pratama Muh, Rifandi Muh, Sanusi W, Side S. Stability analysis and numerical simulation of SEIR model for pandemic COVID-19 spread in Indonesia. Chaos, Solitons & Fractals. 2020;139:110072.

4. Djalante R, Lassa J, Setiamarga D, Sudjatma A, Indrawan M, Haryanto B, et al. Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020. Progress in Disaster Science. 2020;6:100091.

5. Sahoo BK, Sapra BK. A data driven epidemic model to analyse the lockdown effect and predict the course of COVID-19 progress in India. Chaos, Solitons & Fractals. 2020;139:110034.

6. https://covid19.who.int/ [Internet]. Available from: https://covid19.who.int/

7. https://covid19.go.id/ [Internet]. Available from: https://covid19.go.id/

8. Oum TH, Wang K. Socially optimal lockdown and travel restrictions for fighting communicable virus including COVID-19. Transport Policy. 2020;96:94–100.

9. Ng WL. To lockdown? When to peak? Will there be an end? A macroeconomic analysis on COVID-19 epidemic in the United States. Journal of Macroeconomics. 2020;65:103230.

10. Lalwani S, Sahni G, Mewara B, Kumar R. Predicting optimal lockdown period with parametric approach using three-phase maturation SIRD model for COVID-19 pandemic. Chaos, Solitons & Fractals. 2020;138:109939.

11. Sardar T, Nadim SS, Rana S, Chattopadhyay J. Assessment of Lockdown Effect in Some States and Overall India: A Predictive Mathematical Study on COVID-19 Outbreak. Chaos, Solitons & Fractals. 2020;110078.

12. Ghosal S, Bhattacharyya R, Majumder M. Impact of complete lockdown on total infection and death rates: A hierarchical cluster analysis. Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 2020;14:707–11.

13. Atalan A. Is the lockdown important to prevent the COVID-19 pandemic? Effects on psychology, environment and economy-perspective. Annals of Medicine and Surgery. 2020;56:38–42.

14. Ocampo L, Yamagishi K. Modeling the lockdown relaxation protocols of the Philippine government in response to the COVID-19 pandemic: An intuitionistic fuzzy DEMATEL analysis. Socio-Economic Planning Sciences. 2020;100911.

15. Buso IM, De Caprariis S, Di Cagno D, Ferrari L, Larocca V, Marazzi F, et al. The effects of COVID-19 lockdown on fairness and cooperation: Evidence from a lablike experiment. Economics Letters. 2020;196:109577.

16. Vinceti M, Filippini T, Rothman KJ, Ferrari F, Goffi A, Maffeis G, et al. Lockdown timing and efficacy in controlling COVID-19 using mobile phone tracking. EClinicalMedicine. 2020;100457.

17. Sangiorgio V, Parisi F. A multicriteria approach for risk assessment of Covid-19 in urban district lockdown. Safety Science. 2020;130:104862.

18. Acharya R, Porwal A. A vulnerability index for the management of and response to the COVID-19 epidemic in India: an ecological study. The Lancet Global Health. 2020;8:e1142–51.

19. Lin Y, Hu X, Li X, Wu X. Mining stable patterns in multiple correlated databases. Decision Support Systems. 2013;56:202–10.

20. Nti IK, Adekoya AF, Weyori BA. A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction. J Big Data. 2021;8:17.

21. Ait-Mlouk A, Agouti T, Gharnati F. Mining and prioritization of association rules for big data: multi-criteria decision analysis approach. J Big Data. 2017;4:42.

22. Killeen GF, Kiware SS. Why lockdown? Why national unity? Why global solidarity? Simplified arithmetic tools for decision-makers, health professionals, journalists and the general public to explore containment options for the 2019 novel coronavirus. Infectious Disease Modelling. 2020;5:442–58.

23. Saaty TL, Vargas LG. Decision Making with the Analytic Network Process [Internet]. Boston, MA: Springer US; 2013 [cited 2021 Jan 8]. Available from: http://link.springer.com/10.1007/978-1-4614-7279-7

24. Saaty TL, Vargas LG. Decision making with the analytic network process: economic, political, social and technological applications with benefits, opportunities, costs and risks. New York: Springer; 2006.

25. Farman H, Jan B, Talha M, Zar A, Javed H, Khan M, et al. Multicriteria-Based Location Privacy Preservation in Vehicular Ad Hoc Networks. Complexity. 2018;2018:1–12.

26. Lin W. Efficient Adaptive-Support Association Rule Mining for Recommender Systems. Data Mining and Knowledge Discovery. 2002;6:83–105.

27. Osadchiy T, Poliakov I, Olivier P, Rowland M, Foster E. Recommender system based on pairwise association rules. Expert Systems with Applications. 2019;115:535–42.

28. Jooa J, Bangb S, Parka G. Implementation of a Recommendation System Using Association Rules and Collaborative Filtering. Procedia Computer Science. 2016;91:944–52.

29. Paranjape-Voditel P, Deshpande U. A stock market portfolio recommender system based on association rule mining. Applied Soft Computing. 2013;13:1055–63.

30. Kim YS, Yum B-J. Recommender system based on click stream data using association rule mining. Expert Systems with Applications. 2011;38:13320–7.

31. Nassar N, Jafar A, Rahhal Y. Multi-criteria collaborative filtering recommender by fusing deep neural network and matrix factorization. J Big Data. 2020;7:34.

32. Walek B, Fojtik V. A hybrid recommender system for recommending relevant movies using an expert system. Expert Systems with Applications. 2020;158:113452.

33. Hikmawati E, Maulidevi NU, Surendro K. A multi-criteria recommender system model for determining lockdown decision of Covid-19 cases in Indonesia. Proceedings of the 2021 10th International Conference on Software and Computer Applications. submited. p. 9.

34. Hikmawati E, Maulidevi NU, Surendro K. Adaptive rule: A novel framework for recommender system. ICT Express. 2020;S2405959520300916.

35. Gera J, Kaur H. A novel framework to improve the performance of crowdfunding platforms. ICT Express. 2018;4:55–62.

36. Farman H, Javed H, Jan B, Ahmad J, Ali S, Khalil FN, et al. Analytical network process based optimum cluster head selection in wireless sensor network. Shi Y, editor. PLoS ONE. 2017;12:e0180848.

37. El Mazouri FZ, Abounaima MC, Zenkouar K. Data mining combined to the multicriteria decision analysis for the improvement of road safety: case of France. Journal of Big Data [Internet]. 2019 [cited 2019 Sep 25];6. Available from: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-018-0165-0

38. Ait-Mlouk A, Gharnati F, Agouti T. An improved approach for association rule mining using a multi-criteria decision support system: a case study in road safety. Eur Transp Res Rev. 2017;9:40.

39. Sahafizadeh E, Sartoli S. Epidemic curve and reproduction number of COVID-19 in Iran. Journal of Travel Medicine. 2020;27:taaa077.

40. Pikobar [Internet]. Available from: https://pikobar.jabarprov.go.id/data

41. Profil Kesehatan Jawa Barat [Internet]. Available from: http://diskes.jabarprov.go.id/

42. Badan Pusat Statistik [Internet]. Available from: https://www.bps.go.id/

43. Ye G-Y, Xu K-J, Wu W-K. Standard deviation based acoustic emission signal analysis for detecting valve internal leakage. Sensors and Actuators A: Physical. 2018;283:340–7.

44. Zhao X, Jiang N, Liu J, Yu D, Chang J. Short-term average wind speed and turbulent standard deviation forecasts based on one-dimensional convolutional neural network and the integrate method for probabilistic framework. Energy Conversion and Management. 2020;203:112239.

Figures

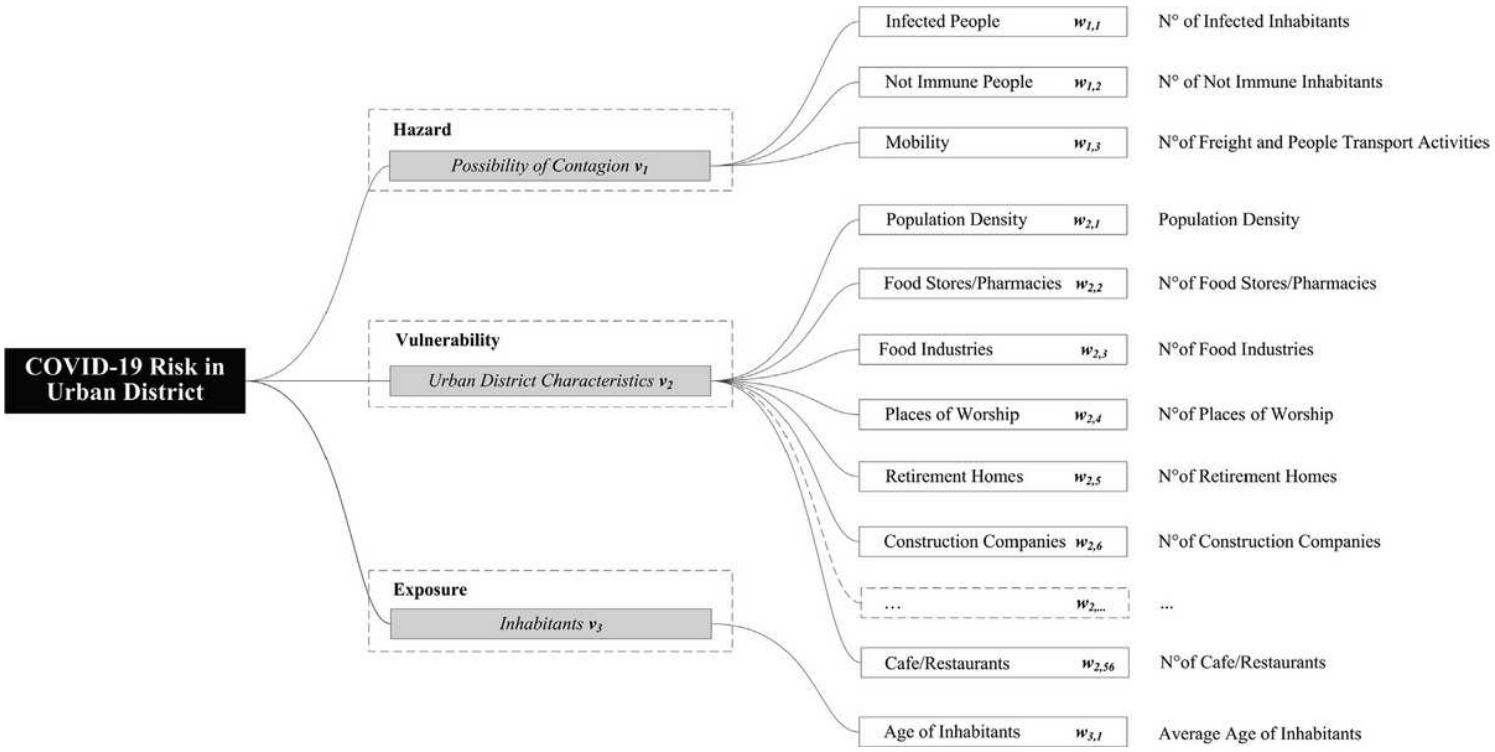

Figure 1

Structure of the problem in Criteria, Sub-Criteria and Intensity [17]

	Variable description	Source
Socioeconomic		
Scheduled tribe or caste households	Calculated as proportion of households belonging to scheduled caste or tribe	National Family Health Survey-4 (Household file), India (2015–16)[6]
Education level in population	Calculated as proportion of population who completed secondary or higher level of education	National Family Health Survey-4 (Person file), India (2015–16)[6]
Poor households	An asset deprivation indicator was computed as the proportion of households that did not have any of the following: a motorised vehicle (a two-wheeler, car or truck, or tractor), television, computer, bicycle, refrigerator, thresher, or air-conditioner or cooler	National Family Health Survey-4 (Household file), India (2015–16)[6]
Demographic		
Elderly population	Calculated as proportion of individuals in the population aged 60 years or older	National Family Health Survey-4 (Person file), India (2015–16)[6]
Urbanisation	Calculated as proportion of urban households among all households	National Family Health Survey-4 (Household file), India (2015–16)[6]
Population density	Calculated as a ratio of population of a unit (district or state) and its area in km^2	Area data: 2011 census;[19] population data: linearly projected population for 2019 using growth rate calculated for each district based on 2001 and 2011 census
Housing and hygiene condition		
People per room	Calculated as the mean number of people residing per room used for sleeping in a household	National Family Health Survey-4 (Household file), India (2015–16)[6]
Households with no toilet facility	Calculated as proportion of households reporting no availability of toilet facility within premises	National Family Health Survey-4 (Household file), India (2015–16)[6]
Households with no hand-hygiene facility	Calculated as percent households with no availability of water and soap or detergent at place of handwashing	National Family Health Survey-4 (Household file), India (2015–16)[6]
Availability of health care		
Households with health insurance	Calculated as proportion of households with at least one member covered under any health insurance scheme	National Family Health Survey-4 (Household file), India (2015–16)[6]
Households without easy access to public health facility	Calculated as proportion of households reported having no nearby public health facility	National Family Health Survey-4 (Household file), India (2015–16)[6]
Availability of public hospitals (at district level)	Calculated as number of public hospitals (primary health centre and above) per 100 000 population	Rural health statistics 2018 and linearly projected population for 2019 using growth rate calculated for each district based on 2001 and 2011 census[20]
Availability of hospital beds (at state level)	Calculated as number of public or private hospital beds per 1000 population	National health profile 2019[21]
Epidemiological		
Men with any chronic morbidity	Calculated as proportion of men aged 40–54 years with chronic health conditions, such as cardiovascular disease, diabetes, asthma, or cancer	National Family Health Survey-4 (Men's file), 2015–16[6]
Men who smoke	Calculated as proportion of men who smoke tobacco	National Family Health Survey-4 (Men's file), 2015–16[6]
Women with any chronic morbidity	Calculated as proportion of women aged 40–49 years with chronic health conditions, such as cardiovascular disease, diabetes, asthma, or cancer	National Family Health Survey-4 (Women's file), 2015–16[6]

Figure 2

Domains of vulnerability and variables within [18]

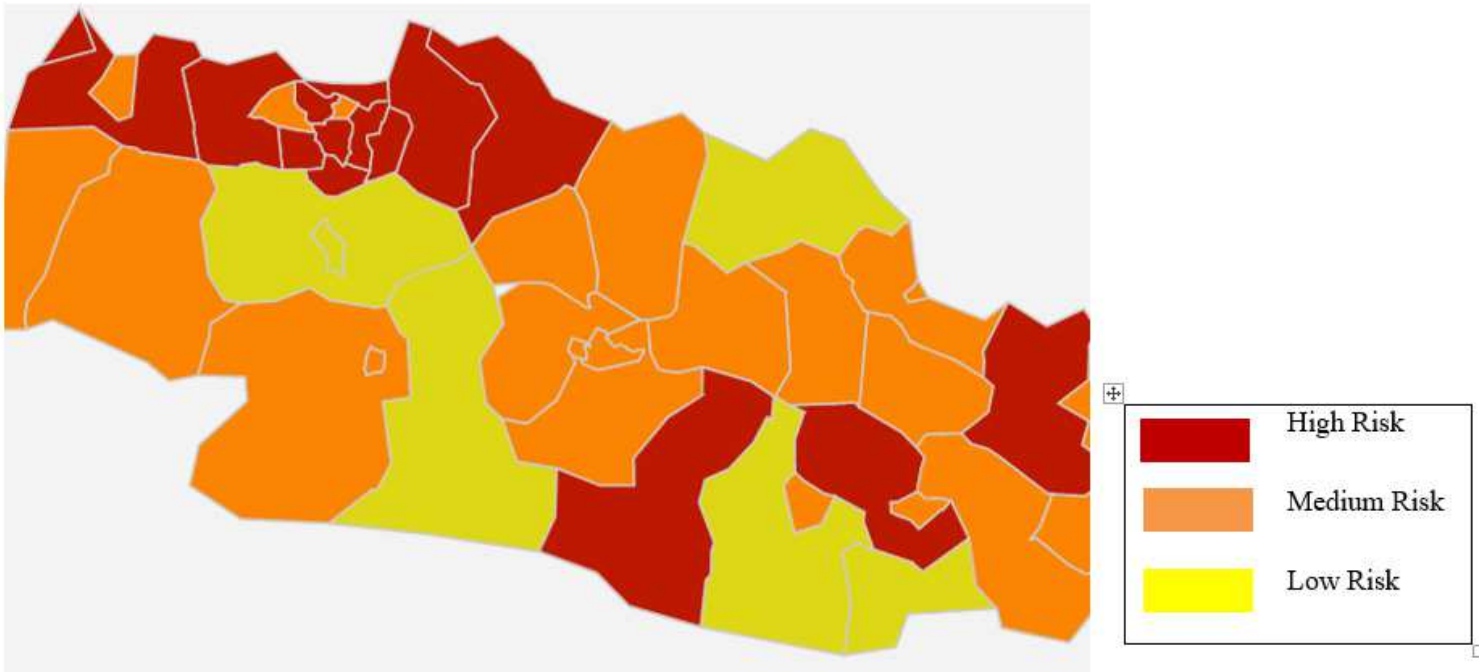

Figure 3

Covid-19 risk map for the West Java region [7] Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

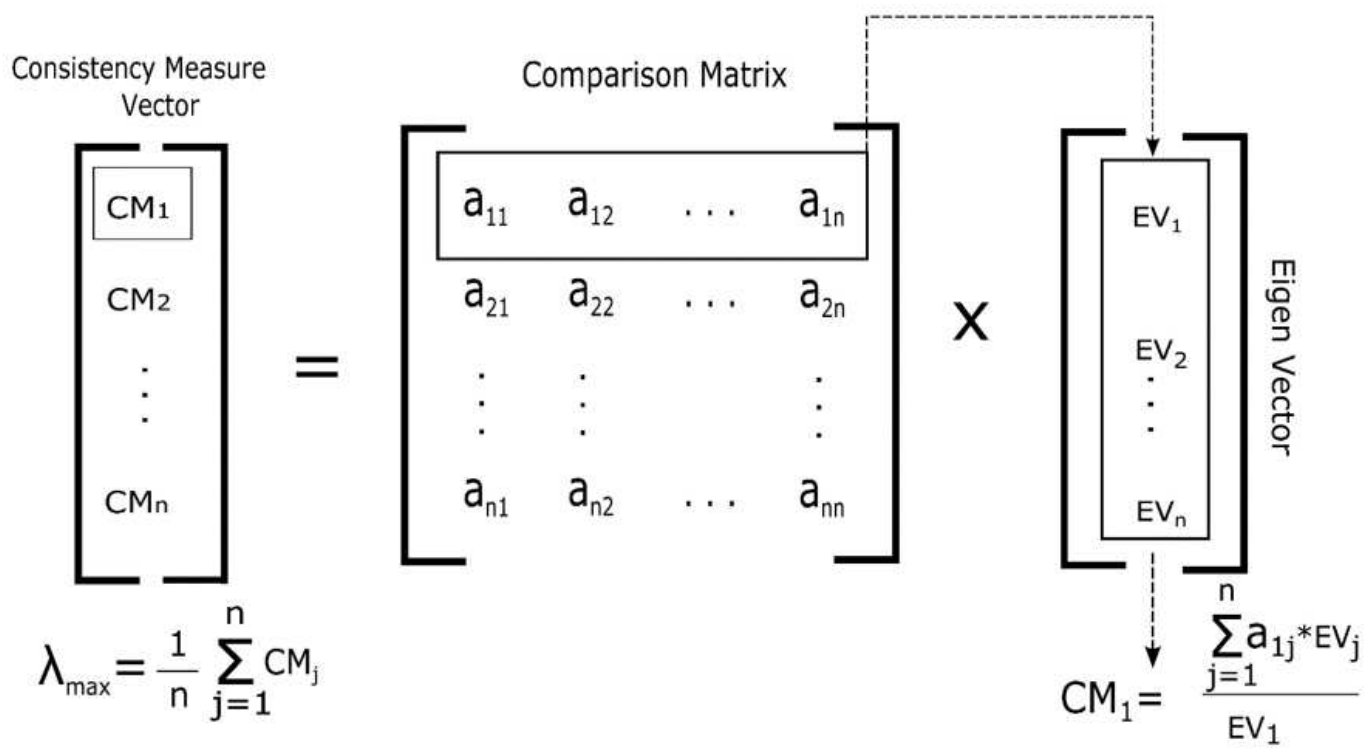

Figure 4

Consistency Measure calculation (CM) [36]

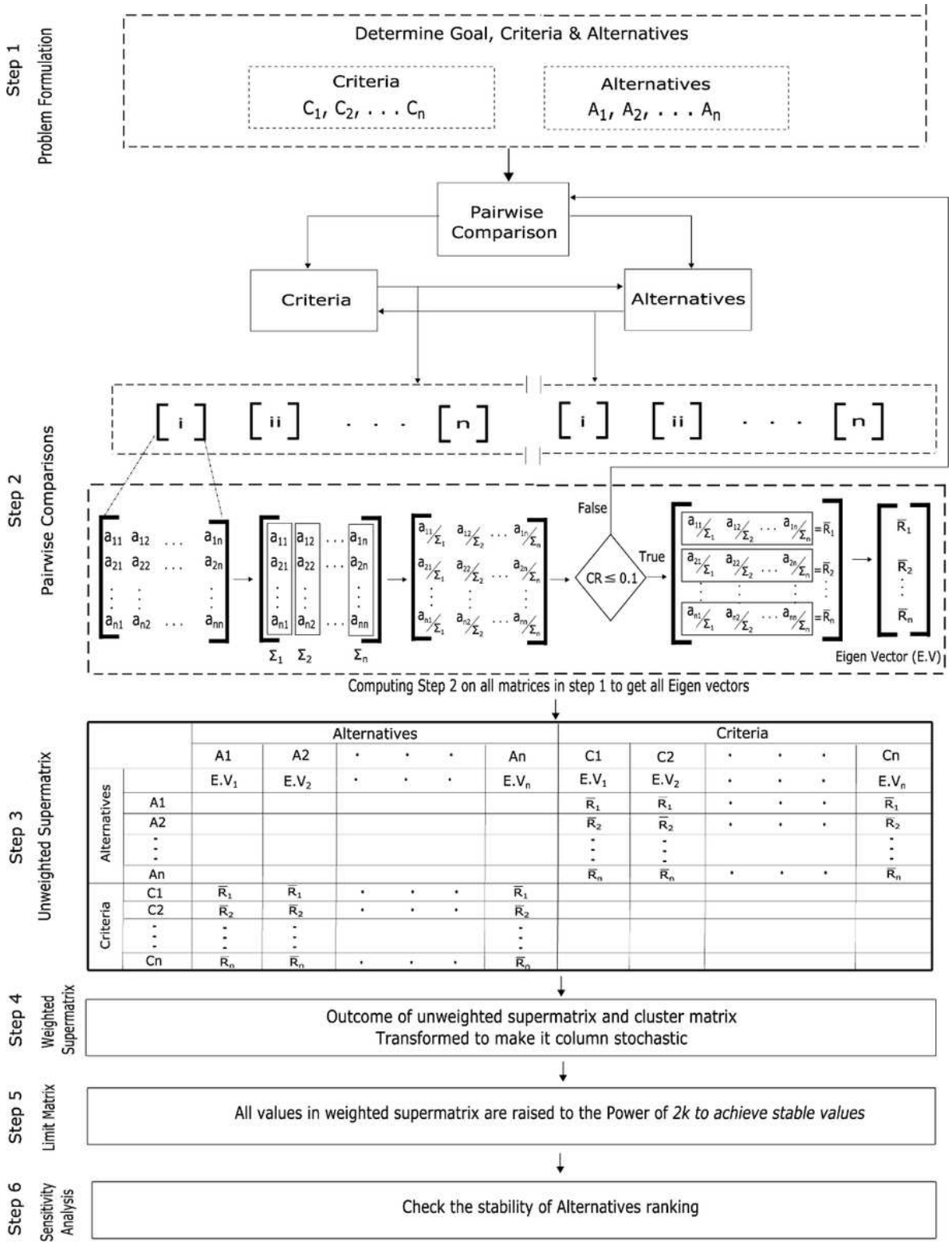

Figure 5

General Step for ANP [36]

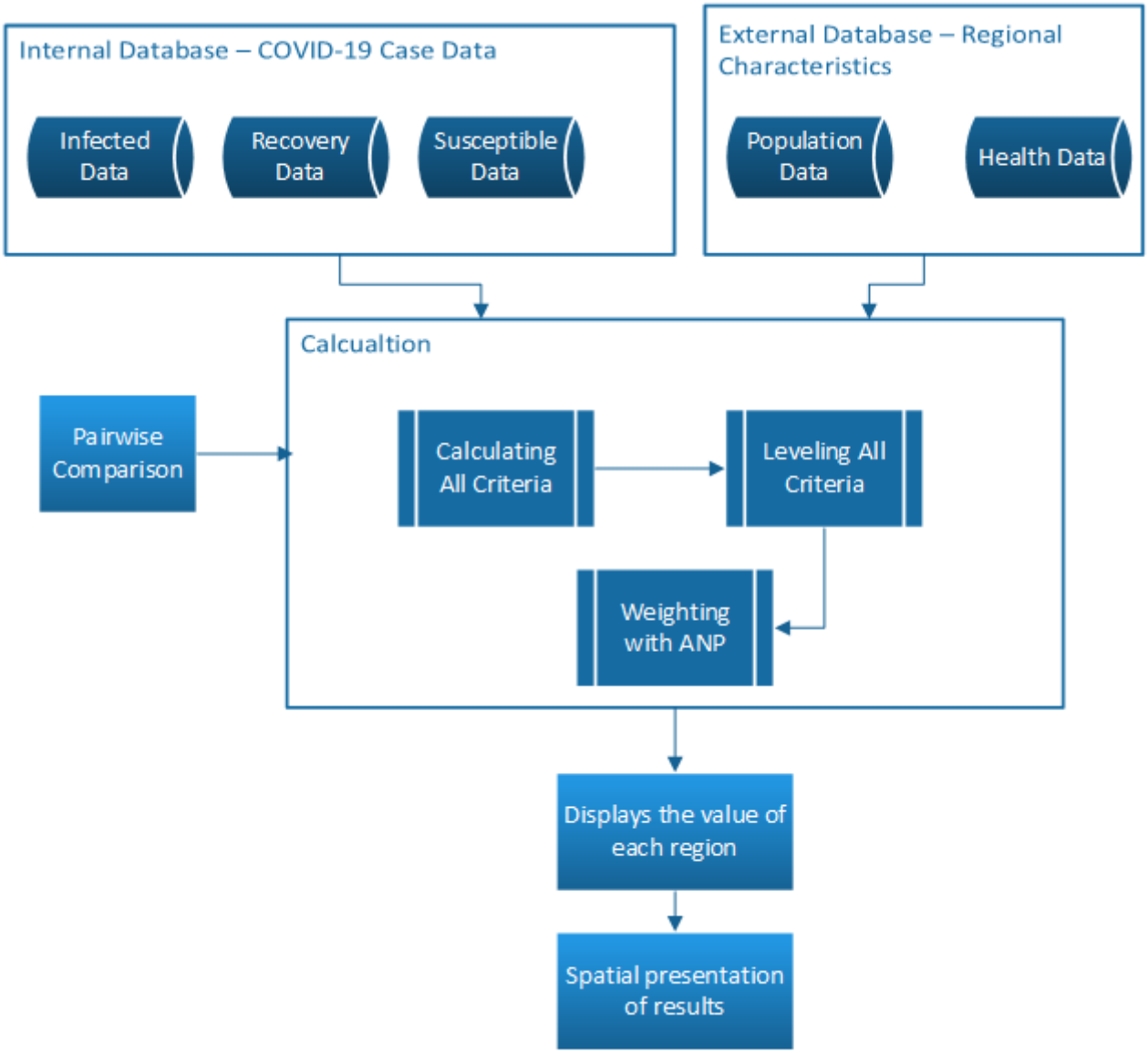

Figure 6

Model Multicriteria Recommender System

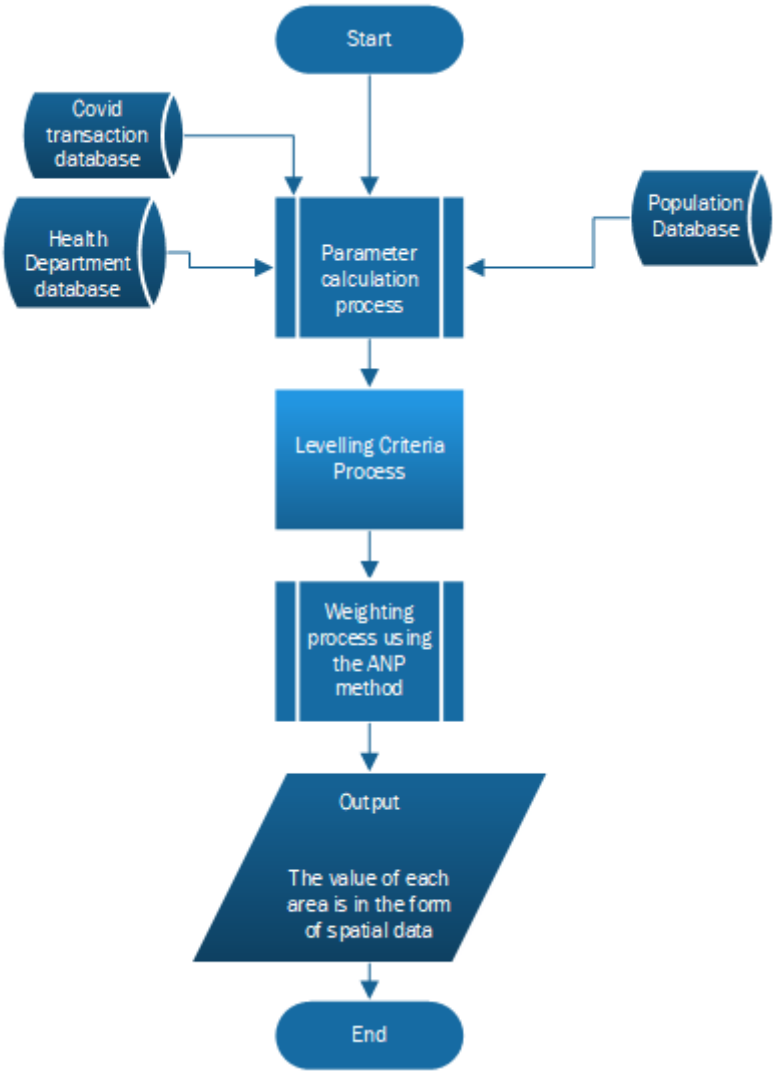

Figure 7

Main Flow of Multicriteria Recommender System Model

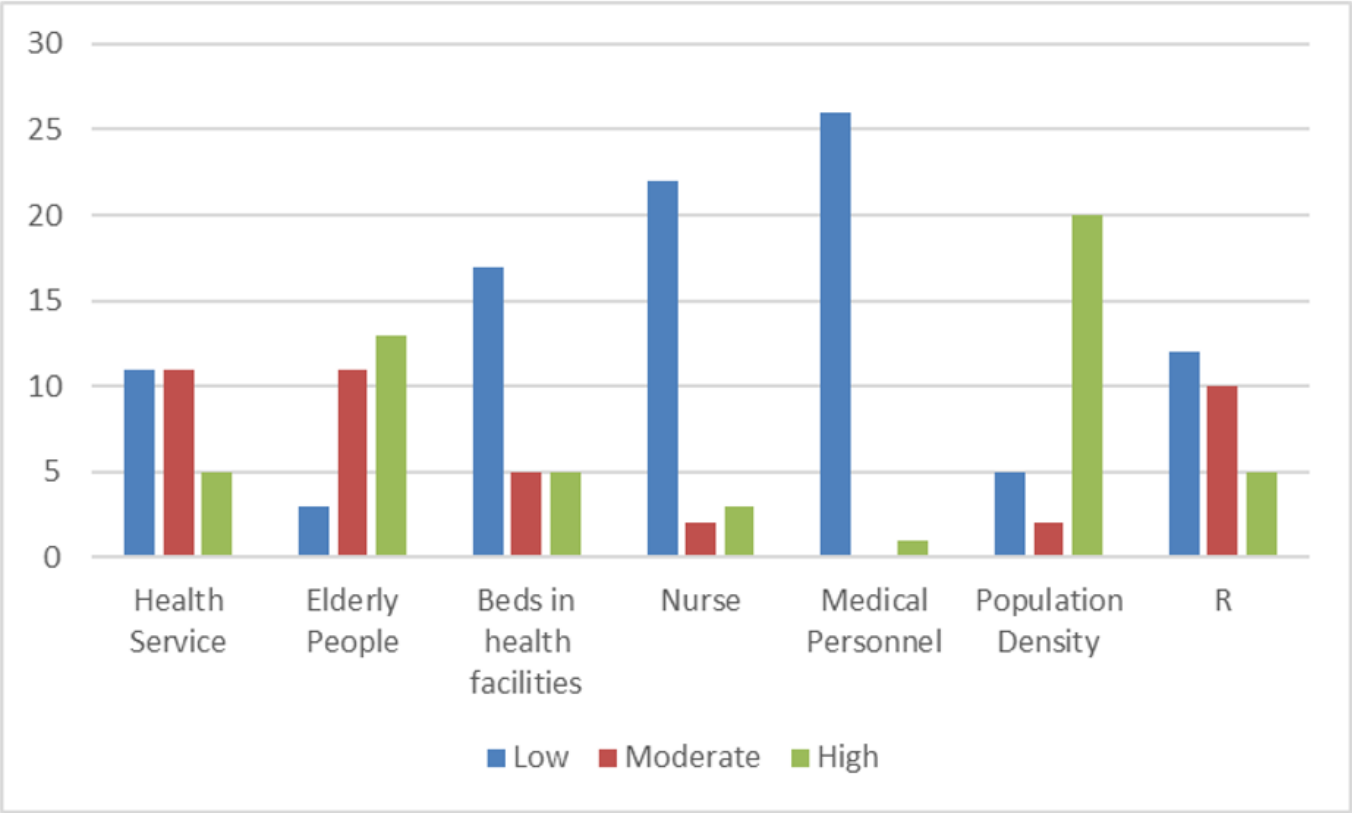

Figure 8

Regional Characteristics in West Java

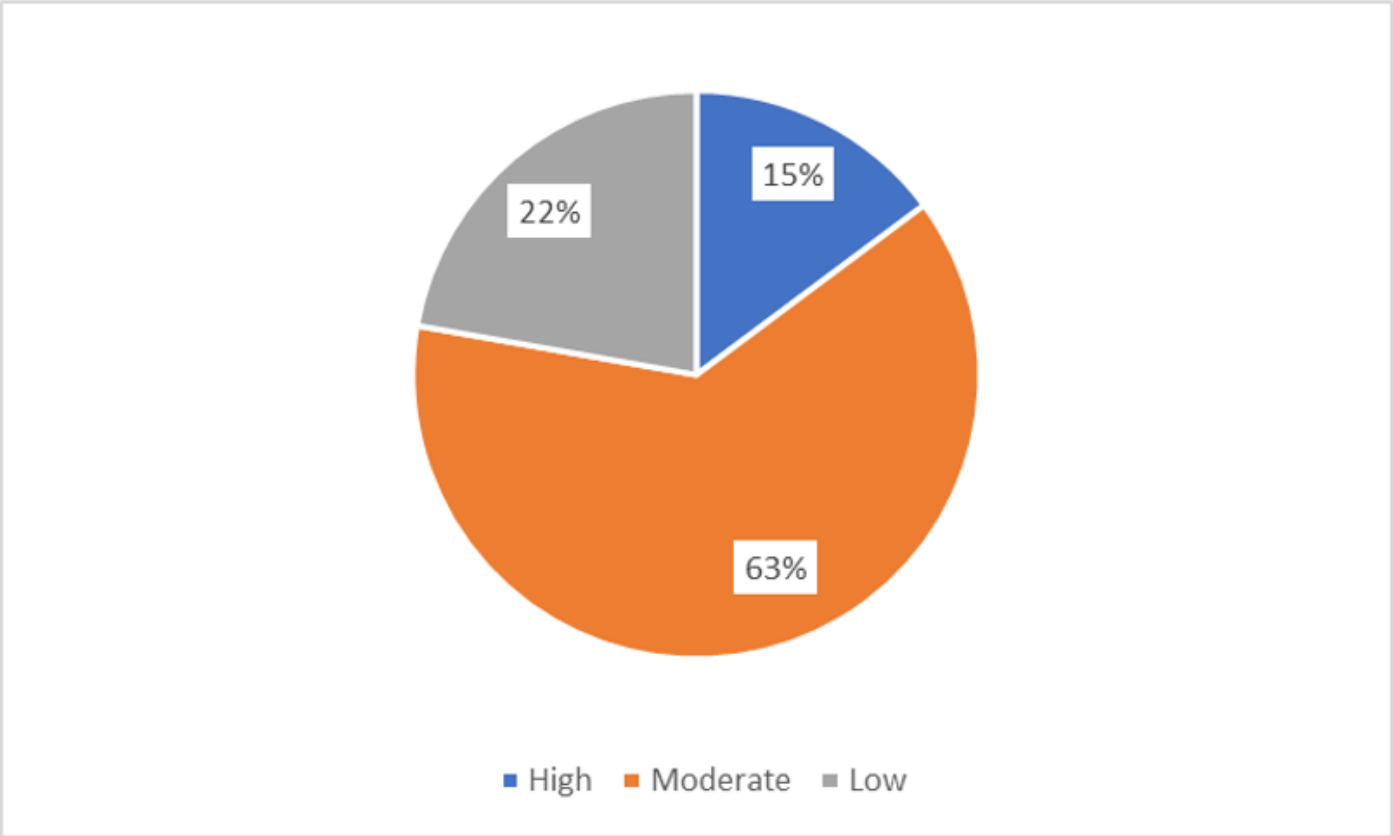

Figure 9

Percentage of area assessment results in West Java

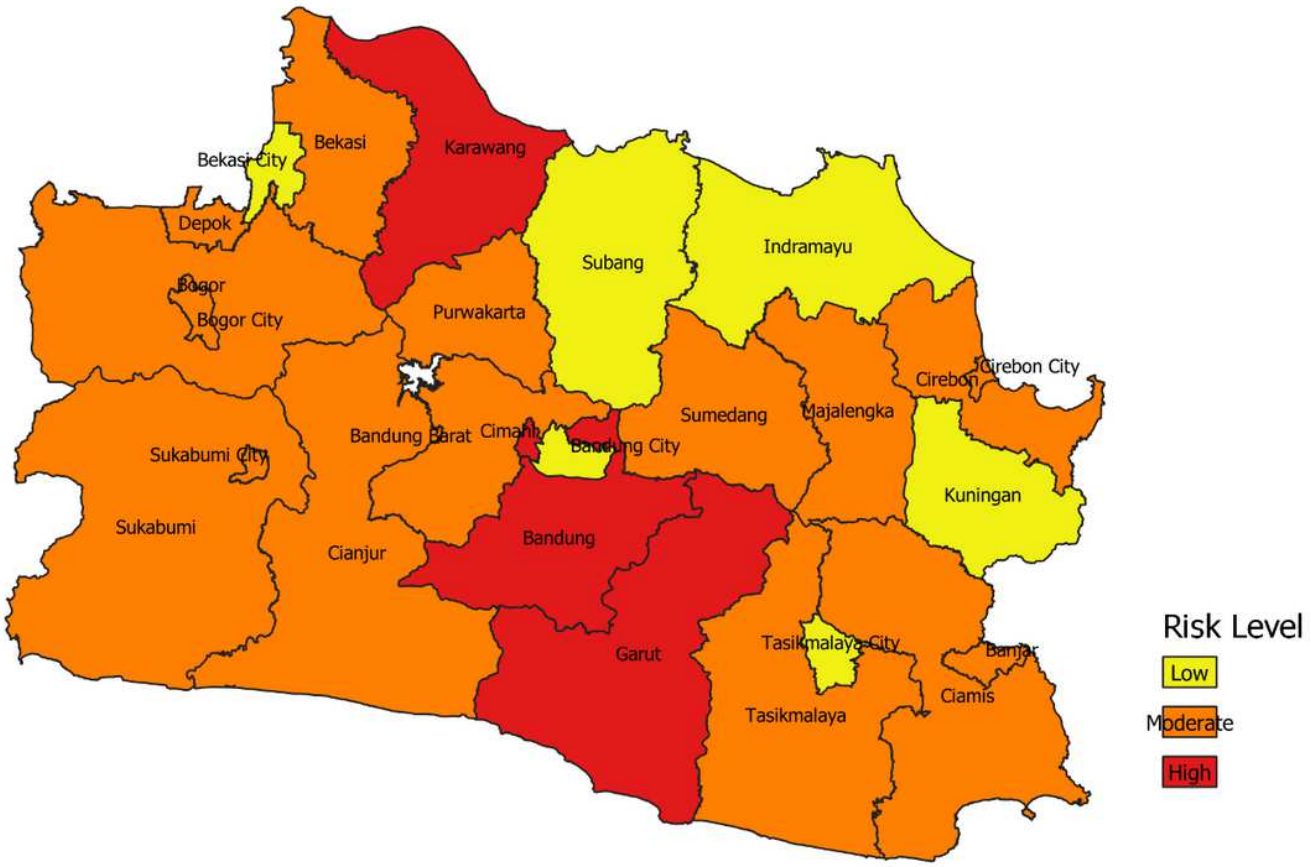

Figure 10

Spatial visualization of the final result Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.