JOIV : Int. J. Inform. Visualization, 8(3) - September 2024 1445-1452

INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Distribution Model of Personal Protective Equipment (PPE) Using the Spatial Dominance Test and Decision Tree Algorithm

Vega Purwayoga [a,*], Siti Yuliyanti [a], Andi Nurkholis [b], Harry Gunawan [c], Sokid [c], Nuri Kartini [d]

[a] *Department of Informatics, Faculty of Engineering, Siliwangi University, Tasikamalaya, Indonesia*
[b] *Department of Informatics, Faculty of Engineering and Computer Sciences, Teknokrat Indonesia University, Lampung, Indonesia*
[c] *Department of Informatics Engineering, Faculty of Engineering, Universitas Muhammadiyah Cirebon, Cirebon, Indonesia*
[d] *Department of Industrial Engineering, Faculty of Engineering, Universitas Muhammadiyah Cirebon, Cirebon, Indonesia*

*Corresponding author: *vega.purwayoga@unsil.ac.id*

Abstract*—The COVID-19 case has developed positively, but preventive measures must be taken to anticipate SARS-CoV-2 mutations. Anticipation can include policies, preparing health workers, and providing personal protective equipment. Personal Protective Equipment (PPE) availability is a big challenge in handling pandemics, especially COVID-19. The level of need for PPE in an area depends on the number of COVID-19 cases. This research provides a solution to overcome the availability of PPE by applying the concept of cross-regional collaboration. Areas with low COVID-19 case rates can help areas with high COVID-19 case rates by sending PPE assistance. Implementing the cross-regional collaboration concept is assisted by the spatial dominance test algorithm, namely the spatial skyline query. Spatial Skyline Query works by searching for the most ideal area. The ideal area is an area with low COVID-19 case criteria. The low number of positive cases, death cases, probable cases, and close contact cases supports the low number of COVID-19 cases. Areas with the highest number of recovered cases are also priorities. The SSQ model was developed into two models for searching priority areas for PPE assistants. The first model is Sort Filter Skyline 1 (SFS1), and the second is Sort Filter Skyline 2 (SFS2). SFS1 is a form of SFS algorithm optimization that searches for the best 50% of all regions. SFS2 modifies SFS1 by selecting areas whose distance is <= the average distance of the area to the Health Crisis Centre of the Ministry of Health of the Republic of Indonesia. This research involves searching for priority areas and applying a prediction algorithm to extract knowledge built from the prediction model. The algorithm used is C5.0. The data used to apply the prediction algorithm results from the application of SFS1 and SFS2. The results of testing the prediction model by the C5.0 algorithm produced an accuracy of 77.26% for SFS1 data and 92.01% for SFS2. The average rules resulting from the C5.0 algorithm are three for SFS1 and two for SFS2.

***Keywords*—COVID-19; cross-regional collaboration; distribution; personal protective equipment; spatial skyline query.**

Manuscript received 9 Jan. 2024; revised 27 Mar. 2024; accepted 15 Apr. 2024. Date of publication 30 Sep. 2024.

I. INTRODUCTION

COVID-19 is an epidemic that attacks the respiratory system [1]. The first COVID-19 case that appeared in Indonesia occurred in March 2020. COVID-19 cases increased very quickly; it was recorded that in July 2020, the number of positive instances reached 76,000 cases, and 3,500 people were reported to have died [2]. The increase in cases is motivated by human behavior interacting with each other, so there is a need for policies to reduce the spread of the COVID-19 outbreak [3]. In June 2023, COVID-19 cases experienced a positive trend, but it is still possible for a spike in cases to occur due to SARS-CoV-2 mutations [4].

Several policies have been implemented to reduce the spread of the COVID-19 outbreak, such as maintaining distance and using masks [5]. Maintaining distance and wearing a mask when doing activities are practical ways that are recommended to reduce the spread of COVID-19 cases [6]. Not only do people need masks, but health workers also need masks for personal protection. Health workers are among the most important parts of handling COVID-19, so personal protective equipment (PPE) is also essential for health workers. The PPE needed for health workers includes respirators, gloves, gowns and eye protection [7].

The high number of COVID-19 cases affects the need for PPE in hospitals [8], [9]. Research [9] shows that the availability of PPE is still uneven. Many hospitals still lack PPE, indicating that the distribution of PPE is not yet ideal. The availability of PPE needs to be calculated based on the number of daily COVID-19 cases, health workers, and patients [10].

Research related to the distribution of PPE has been carried out [2] by measuring regions based on the level of COVID-19 cases that occur in an area. Research [2], [11] applies a dominance testing algorithm, namely spatial skyline query (SSQ). SSQ has succeeded in recommending priority areas for PPE recipients based on the dominance test. Areas with high COVID-19 cases are priority areas for receiving PPE assistance. Research [12] and [13] shows that areas with low COVID-19 cases hope to be able to help areas with high COVID-19 cases. Research [2] applies the concept of cross-regional collaboration, where every individual, organization, and other entity works together to handle COVID-19. Cross-regional collaboration is an effective solution to help with resource management in response to handling COVID-19 [14], [15]. Research [16] states that the PPE procurement process is one of the significant challenges in handling COVID-19. In research, [16] created a distribution model by measuring the priority of PPE needs in an area. A critical factor in the process of distributing goods is distance. Distance determines costs in the process of distributing goods so that the closest distance is the priority [17]. Research [10] has predicted the need for PPE in a hospital when handling COVID-19. The results of the research [10] can be applied to predict the need for PPE, but this research only predicts the need for PPE and has not yet reached the stage of developing a model for the distribution of PPE. Several previous studies have revealed that SSQ is one of the practical algorithms in searching for dominant objects based on the preferences held [18].

Several previous studies have measured the need for PPE, and some of them have applied several essential aspects, one of which is the spatial aspect, namely distance [2]. Distance is essential when developing a model with the concept of cross-regional collaboration. Research [10] has predicted the need for PPE based on COVID-19 cases in a hospital. However, no previous research has developed a prediction model for priority areas for PPE distributors. Prediction models can become knowledge in supporting decisions based on past events [19]. Based on research [19], [20], one of the prediction models that perform well is decision trees. Where a decision tree produces knowledge in the form of a decision tree and a group of rules, one of the popular decision tree algorithms is C5.0 [21]. The reason the algorithm uses C5.0 is because it is easy for humans to understand [22]. Another reason for using the C5.0 algorithm lies in its performance aspect, as the C5.0 algorithm has been demonstrated to be effective in predicting disasters [23]. Research [24] compared several machine learning algorithms for prediction with the best results achieved by the C5.0 algorithm. This research applies the concept of cross-regional collaboration to distribute PPE in handling COVID-19. This research applies a dominance testing algorithm to find priority areas for PPE distributors to implement the cross-regional collaboration concept. This study improves upon prior research [2] by augmenting it with the C5.0 algorithm for knowledge extraction and prediction of areas serving as PPE distributors. Based on several previous studies, the SSQ algorithm has shown good performance in recommending specific objects, particularly in PPE distribution. In addition to implementing the C5.0 algorithm, this research also optimizes the SSQ algorithm by selecting regions with the minimum distance.

II. Material and Method

A. Data and Study Area

This research's data and study area are daily COVID-19 case data from West Java Province, Indonesia. The dataset consists of tabular data and spatial data/shapefiles, consisting of regional administrative boundaries (districts/cities). The daily case data used starts from August 2020 to July 2021, a period of high COVID-19 cases. The daily cases obtained in that period were 5712 data from 27 regions in West Java Province. Detailed information regarding the data used is presented in Table 1.

This research has several stages, as can be seen in Figure 1. The stages are data collection, data preprocessing on tabular and spatial data, searching for priority areas using spatial dominance testing (labeling), and classification using the C5.0 algorithm.

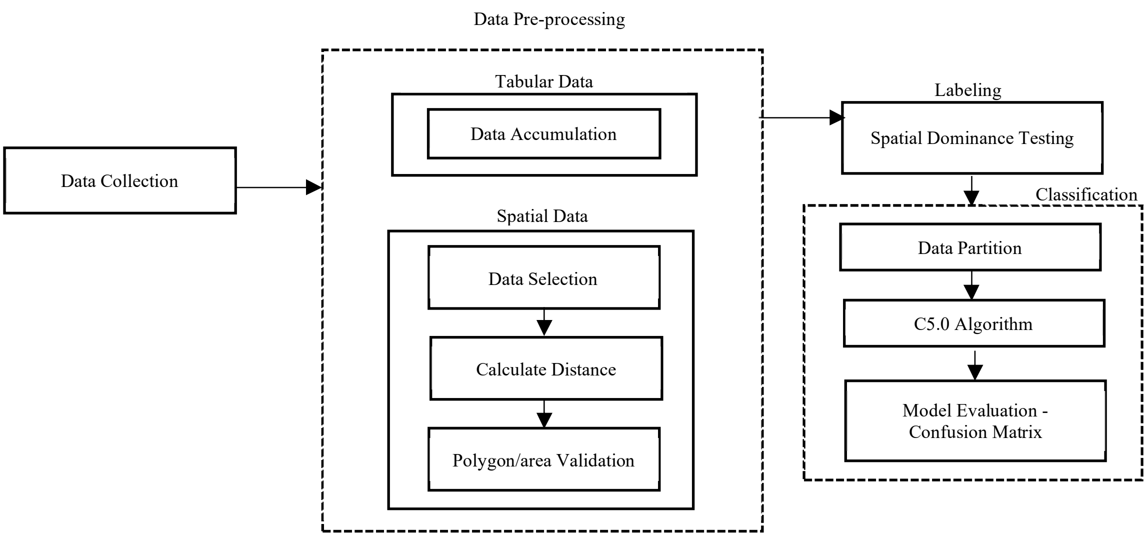

Fig. 1 Research stages

TABLE I
DATASET

Attributes name	Description
Date	The date of occurrence of the COVID-19 case
District	The name of the district/city
Close Contact	People who interact with people who are confirmed positive for COVID-19
Probable	Has symptoms of severe acute respiratory tract infection
Positive	People who have been declared positive for Coronavirus infection based on the results of a laboratory examination in the form of PCR
Recovered	People who have tested negative for COVID-19 after previously being positive
Death	condition when a person who is in the probable or confirmed COVID-19 category dies.

B. Data Collection and Pre-processing

Data on COVID-19 cases was obtained from the West Java COVID-19 Information and Coordination Center (pikobar.jabar.prov.id), the official website of the COVID-19 task force for West Java Province, Indonesia. Spatial data was obtained from the Geospatial Information Agency (BIG). Spatial data contains district/city-level polygons in Indonesia.

The data preprocessing stages for tabular data are accumulating daily case data into monthly data and getting COVID-19 incidents for each district/city. The data accumulation process is assisted by using the tdyverse library. Tidyverse contains a set of libraries that are usually used to manipulate data. The preprocessing stage is selected in spatial data by taking polygons or districts/cities in West Java Province. The district/city selection process is assisted using the QGIS application. Check whether the district/city polygon is valid or not. When a polygon is invalid, it is processed with the fix geometry's function. The advanced data preprocessing stage for spatial data is measuring the distance between the midpoint of the polygon of each district/city and the Health Crisis Center of the Ministry of Health of the Republic of Indonesia. Distance measurements use Haversine Distance [25] as presented in Formula 1. Haversine distance is one of the best algorithms for measuring distance because it considers the earth's curve [26].

$$d = \sqrt{2r.\arcsin\left(sin^2\left(\frac{\varphi_2-\varphi_1}{2}\right) + \cos(\varphi_1).\cos(\varphi_2).sin^2\left(\frac{\lambda_2-\lambda_1}{2}\right)\right)} \quad (1)$$

C. Spatial Sort Filter Skyline (Spatial SFS)

Skyline query is a multi-objective query that aims to find the most suitable object based on user preferences [28], [29]. The data labeling process in this research utilizes a recommendation algorithm, Sort Filter Skyline (SFS). The SFS algorithm sorts the data first; then, dominance testing is carried out [18]. The SFS flow is expressed in Algorithm 1 [2]. SFS requires a normalization process and entropy calculation, as presented in (2) and (3).

Normalization of values for each attribute $D_{[a_i]}$ is conducted using the min-max normalization method into the range [0,1] in Equation 2, where f represents the normalization result, D denotes the object, $[a_i]$ represents all objects attached to D, $min_{(a)}$ signifies the smallest value and $max_{(a)}$ denotes the highest value for each attribute of the object. Sorting is performed based on the smallest entropy value because minimum (ascending) preferences dominate the preference utilized. The calculation of the entropy value $E(D)$ is obtained from Equation 3.

This research develops two skyline query models, SFS1 and SFS2. SFS1 applies the SFS model by selecting the best 50% of the skyline object search results. Search for the best 50% of skyline objects using the top function in the *rPref* library. SFS2 selects skyline objects by taking skyline objects with a distance = the average distance from the district/city to the Health Crisis Center of the Ministry of Health of the Republic of Indonesia. The hope is to modify the SFS model.

$$f = \frac{D_{[a_i]} - min_{(a)}}{max_{(a)} - min_{(a)}} \quad (2)$$

$$E(D) = \sum_{i=1}^{d} \ln\left(D_{[a_i]} + 1\right) \quad (3)$$

Algorithm 1. Pseudocode Spatial SFS

Input: Dataset D
Output: The Set of Skyline Points of D

```
1: D ← Dataset after preprocessing
2: S ← First dataset D
3: from 1 to D
4:     if ("D is not dominated") then
5:         write (S, D)
6:     else
7:         remove (S, D)
8:     end if
9: end
```

D. Data Partition

Data partitioning aims to divide data into two parts: training data and testing data. Data partitioning was done using 10-fold cross-validation to see the accuracy established by a model. K=10 is the best choice for the number of K because the results are more stable [21]. The ratio of the distribution of training data and testing data is presented in Table II.

TABLE II
DATA PARTITION

Fold on test data	Fold on train data
1	2,3,4,5,6,7,8,9,10
2	1,3,4,5,6,7,8,9,10
3	1,2,4,5,6,7,8,9,10
…	…
8	1,2,3,4,5,6,7,9,10
9	1,2,3,4,5,6,7,8,10
10	1,2,3,4,5,6,7,8,9

E. Classification using Decision Tree C5.0

The C5.0 algorithm is a decision tree model for classifying data, as the C5.0 classification model requires explanatory attributes and target attributes. C5.0 produces a model as a tree and decision rules [21], [22]. C5.0 improves previous decision tree models such as ID3 and C4.5 [21]. C5.0 can handle discrete and continuous attribute types, which in this research mostly use continuous attributes. In the C5.0 decision tree algorithm, attribute selection uses information gain. In the C5.0 algorithm, the attribute selection measure uses information gain and entropy. Entropy is a parameter to measure the diversity of a data set. The more heterogeneous a data set is, the more significant the entropy value. Information gain (Gain(A)) is a measure of the effectiveness of attribute A in classifying data whose highest gain value is selected as the most crucial attribute [22]. The formulas for entropy and gain values are defined in (4), (5), and (6) [27]. Classification is done using three schemes, as shown in Table 2, to compare the best partition results based on classification performance.

$$Entropy(D) = -\sum_i^m p_i log_2 p_i \quad (4)$$

$$Entropy_A(D) = -\sum_i^m \frac{|D_j|}{|D|} \times Entropy(D_j) \quad (5)$$

$$Gain(A) = Entropy(D) - Entropy_A(D) \quad (6)$$

F. Classification Model Evaluation

The model will be tested using pieces of datasets of data. Namely, data from Model SFS1 and Model SFS2, using a confusion matrix. Confusion matrix is an effective technique for measuring the performance of classification models [28]. The confusion matrix has four variables, namely true positive (TP), true negative (TN), false positive (FP), and false negative (FN), as shown in (6) [29], [30]. Accuracy calculations are carried out using formula 7 [31].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

III. Result and Discussion

A. Data Collection and Pre-processing

In tabular data, daily COVID-19 cases are divided into monthly ones. The process of dividing COVID-19 cases is assisted by using the *strptime* function to change the data format to date/month/year, and then the grouping process is carried out using the break function based on the month. Data is grouped into several data frames to see the growth of COVID-19 cases every month. As a result, there are seven new data frames, namely from August 2020 to July 2021. The reason this period was taken is because COVID-19 cases consistently increased every week. In mid-July, the average number of COVID-19 cases per week was 8.000.

Daily cases for each month are accumulated to see the number of cases in each district/city in a particular month. The number of rows before the accumulation process is the number of districts/cities x the number of days in a month. After the accumulation process, the number of data rows became 27 according to the districts/cities in West Java Province. The accumulation process is assisted by using the *tidyverse* library, which runs the group by function for district/city names with the sum function.

Preprocessing on tabular is checking the validity of polygons using the QGIS application. The following process is to measure the distance using the haversine distance formula in the R programming language. Before the distance measurement process, the steps are to find the coordinates of the polygon's center point for each district/city. The *rgeos* library and the *get_centroid* function assist in finding the polygon midpoint. The results of searching for the polygon midpoint can be seen in Table III. The results obtained are the longitude and latitude values for each polygon. The results of measuring the distance from the Health Crisis Center of the Ministry of Health of the Republic of Indonesia to the regional center point can be seen in Figure 2.

TABLE III
Distance Measurement Results

District	Longitude	Latitude
Bandung Regency	107.6110	-7.099763
Bandung Barat Regency	107.4146	-6.896233
Bekasi Regency	107.1207	-6.215149
Bogor Regency	106.7675	-6.559979
Ciamis Regency	108.4319	-7.292083
…	…	…
Cimahi City	107.4395	-6.598452
Cirebon City	107.7322	-6.484194
Depok City	106.7075	-7.076323
Sukabumi City	107.9808	-6.825066
Tasikmalaya City	108.1413	-7.496892

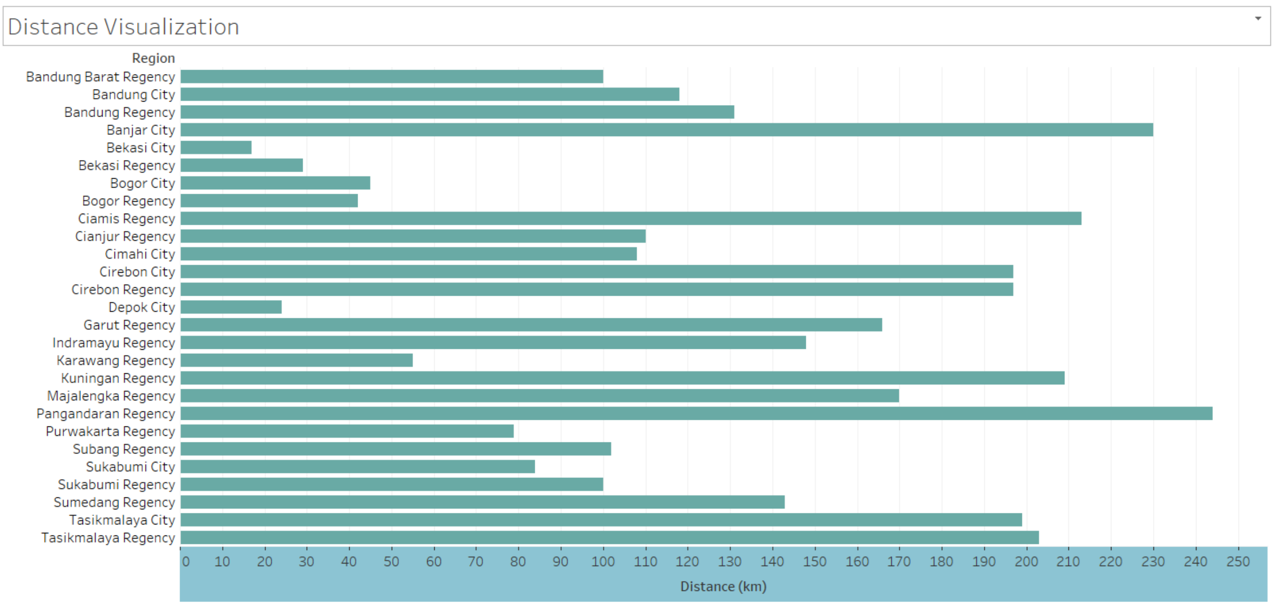

Fig. 2 Distance Visualization

B. *Spatial Sort Filter Skyline (Spatial SFS)*

Two preferences are used: maximum and minimum. The maximum preference function is applied to the recovered attribute, which means that when there are many of them, they are the most recommended; apart from the healing attribute, those with a small number or a short distance are the most recommended. The following process ranks recommendation objects using the skyline query method via the SFS1 and SFS2 algorithms. Normalization of attribute values was carried out using the min-max normalization method into the range [0,1] as presented in Table IV.

TABLE IV
NORMALIZATION RESULTS

District	Close Contact	Probable	Positive	Recovered	Death	Distance
Bandung Regency	0.55957612	0.136842105	0.288135593	0.338680927	0.21428571	0.50220264
Bandung Barat Regency	0.14189713	0.068421053	0.075211864	0.024955437	0.00000000	0.36563877
Bekasi Regency	0.71362109	0.247368421	0.677966102	0.331550802	1.00000000	0.05286344
Bogor Regency	0.00000000	1.000000000	0.277542373	0.124777184	0.35714286	0.11013216
Ciamis Regency	0.33471181	0.015789474	0.009533898	0.007130125	0.00000000	0.86343612
…	…	…	…	…	…	…
Cimahi City	0.96200569	0.005263158	0.085805085	0.035650624	0.00000000	0.40088106
Cirebon City	0.15533730	0.010526316	0.055084746	0.019607843	0.21428571	0.79295154
Depok City	1.00000000	0.168421053	0.616525424	0.483065954	0.85714286	0.03083700
Sukabumi City	0.41922978	0.010526316	0.059322034	0.081996435	0.00000000	0.29515419
Tasikmalaya City	0.17885759	0.268421053	0.012711864	0.000000000	0.00000000	0.80176211

Based on the normalized attribute values, the entropy score of each object is calculated using (3). Data is sorted in ascending order from smallest to most significant entropy. The results of the entropy assessment can be seen in Table V. In contrast to what was done in research [2], in research [2], the sorting was done in descending order. Research [2] looks for areas with high COVID-19 cases so that areas with the most entropy are prioritized. The dominant preference used in research [2] for attributes was the maximum because in research [2], areas with high cases of COVID-19 were priority recipients of PPE assistance.

TABLE V
ENTROPY CALCULATION RESULTS

District	Entropy
Purwakarta Regency	**0.4630604**
Cianjur Regency	0.5091836
Bandung Barat Regency	0.6076621
Subang Regency	0.6741779
Banjar City	0.7158374
…	…
Bandung Regency	1.7186375
Bekasi City	1.8715479
Bekasi Regency	2.3082322
Depok City	2.3726014
Bandung City	2.7301095

The difference between the two methods used is the dominance testing process. SFS1 applies the SFS model to select the best 50% of skyline object search results. Search for the best 50% skyline objects using the top function in the *rPref* library. SFS2 selects skyline objects by taking skyline objects with a distance <= the average distance from the district/city to the Health Crisis Center of the Ministry of Health of the Republic of Indonesia. The recommendation results from the SFS1 and SFS2 models are presented in Table VI. The priority areas for PPE assistants in the SFS2 model are fewer than in the SFS1 model. SFS2 is less because SFS2 re-selects the results from the SFS1 model. Based on Figure 3, spatially adjacent areas will have the same class, namely priority or not priority. Figure 3 shows which areas are priorities for PPE assistants and which are not priorities for PPE assistants. Figures 3 a and 3 b differ in the priority areas. Priority areas are areas with a distance of <= 128 km from the Health Crisis Center of the Ministry of Health of the Republic of Indonesia. Several areas that are a priority in SFS1 are not a priority in SFS2 because the distance does not meet the requirements, such as Pangandaran Regency, Banjar City, Kuningan Regency, Cirebon Regency, Ciamis Regency and Tasikmalaya Regency.

The SFS1 and SFS2 models are in line with research [32], [33]. Objects included in the skyline object can be optimized or selected based on specific preferences to find the best object [32], [33]. Distance optimization carried out in SFS2 is essential because it can minimize distribution time and costs [34].

TABLE VI
RECOMMENDATIONS FOR PRIORITY DISTRICTS

District SFS1	District SFS2
Purwakarta Regency	Purwakarta Regency
Cianjur Regency	Cianjur Regency
Bandung Barat Regency	Bandung Barat Regency
Subang Regency	Sukabumi Regency
Banjar City	Sukabumi City
Sukabumi City	Subang Regency
Tasikmalaya Regency	
Sumedang Regency	
Sukabumi Regency	
Pangandaran Regency	
Kuningan Regency	
Indramayu Regency	
Cirebon Regency	

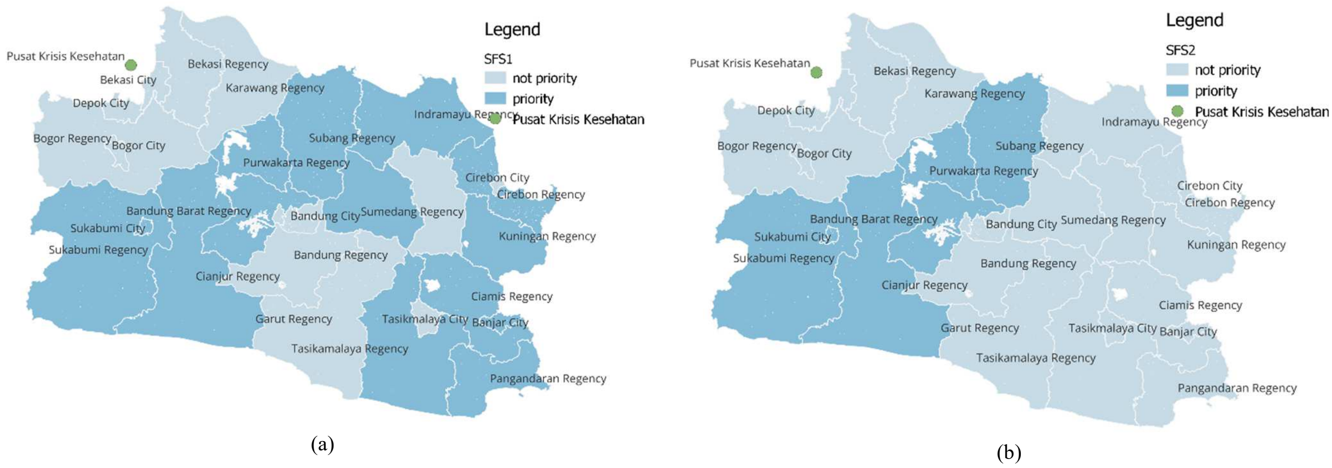

Fig. 3 Visualization of Priority Districts: (a) SFS1 Model and (b) SFS2 Model

C. Data Partition

The labeling results from the SFS model are combined for the classification stage. However, before going through the classification process, some stages need to be carried out, namely data partitioning. Data partitioning in this research uses a 10-fold cross-validation model. Based on research [21] [22]The results of 10-fold cross-validation produce a better and more stable classification model. Data partitioning using 10-fold cross-validation involves dividing the data into ten subsets, where one subset is the testing data and the remaining nine are the training data.

D. Classification Using Decision Tree C5.0

The C5.0 library in the R programming language assists in implementing the C5.0 algorithm. The algorithm is applied to the two data produced by the SFS1 and SFS2 models. C5.0 produces output, namely several rules and attributes that form these rules. The results of developing the C5.0 Algorithm model can be seen in Table VII.

TABLE VII
C5.0 ALGORITHM CLASSIFICATION RESULTS

Data from model	Fold on test data	Fold on train data	Attribute usage	Number of rules
SFS1	1	2,3,4,5,6,7,8,9,10	close contact	2
	2	1,3,4,5,6,7,8,9,10	close contact, death	3
	3	1,2,4,5,6,7,8,9,10	close contact, death	3
	…	…	…	…
	8	1,2,3,4,5,6,7,9,10	close contact	2
	9	1,2,3,4,5,6,7,8,10	close contact	2
	10	1,2,3,4,5,6,7,8,9	close contact, death	3
SFS2	1	2,3,4,5,6,7,8,9,10	district, confirmation	3
	2	1,3,4,5,6,7,8,9,10	district, confirmation	3
	3	1,2,4,5,6,7,8,9,10	district, confirmation	3
	…	…	…	…
	8	1,2,3,4,5,6,7,9,10	District	2
	9	1,2,3,4,5,6,7,8,10	District	3
	10	1,2,3,4,5,6,7,8,9	District	2

The rules generated by the C5.0 Algorithm on SFS1 data are two rules. The rules presented are obtained from Fold 8. Fold 8 produces the best accuracy compared to other Folds. Rules can be shown in a decision tree in Figure 4.

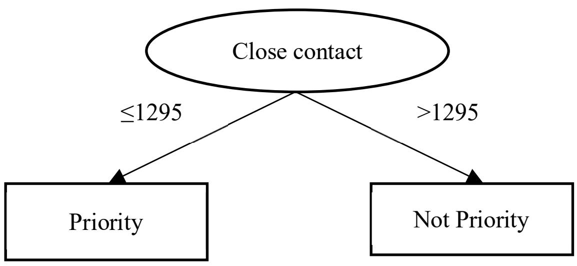

Fig. 4 Decision Trees from SFS1 Data

In the regulations formed from Fold 8, the attribute involved is close contact. The following are the rules created from SFS1 data on Fold 8:

- **IF** close contact > 1295, **THEN** class is not a priority.
- **IF** close contact <= 1295, **THEN** class is a priority.

The rules generated by the C5.0 Algorithm on SFS2 data are 1 rule. The rules presented are obtained from Fold 10. Fold 10 also produces the best accuracy on SFS2 data. Figure 5 shows the rules, which are visualized as a decision tree. In the rules formed from Fold 10, the attribute used is the district's name. The following are the rules created from SFS2 data on Fold 10:

- **IF** district = district A (Bandung City, Bandung Regency, Banjar City, Bekasi City, Bekasi Regency, Bogor City, Bogor Regency, Ciamis Regency, Cimahi Regency, Cirebon City, Cirebon Regency, Depok City, Garut Regency, Indramayu Regency, Karawang Regency, Kuningan Regency, Majalengka Regency, Pangandaran Regency, Sumedang Regency, Tasikmalaya Regency) **THEN** class is non-priority.

- **IF** district = district B (Bandung Barat Regency, Cianjur Regency, Purwakarta Regency, Subang Regency, Sukabumi City, Sukabumi Regency) **THEN** class is a priority.

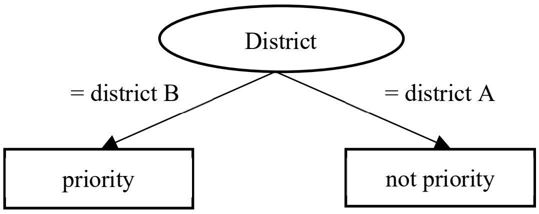

Fig. 5 Decision Trees from SFS2 Data

E. Classification Model Evaluation

Model evaluation is carried out using a confusion matrix. There are two schemes: the first tests the model from SFS1 train with SFS1 test data, and the second tests the model from SFS2 train data with SFS2 test data. The test results have been presented as a confusion matrix, which can be seen in Table VIII. The application of prediction models can be used to estimate the need for assistance in disaster management, especially in the COVID-19 pandemic [35]. The decision tree formed can be used as a model and can support policies that each region can help each other based on regional resilience to a disaster or pandemic [36].

TABLE VIII
CONFUSION MATRIX

Data from model	**Actual Class**	**Prediction Class**	
		Not Priority	Priority
SFS1	Not Priority	7	0
	Priority	2	10
SFS2	Not Priority	11	1
	Priority	0	7

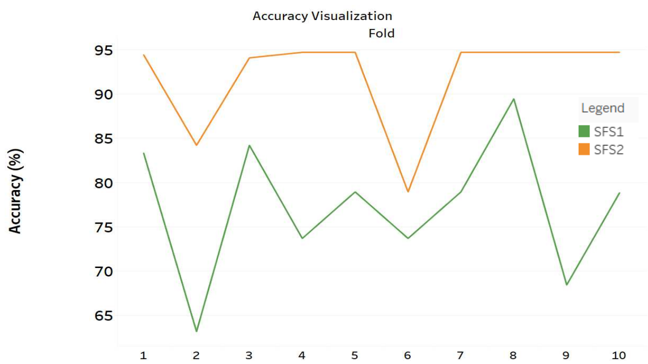

Fig. 6 Accuracy Testing

Figure 6 presents the results of accuracy testing for each data and each Fold. The results of model formation and data testing on SFS2 data produce better accuracy than SFS1 data. As can be seen in Table IX, the average accuracy of SFS1 data is 77.26, while SFS2 is 92.01. The number of rules formed from SFS2 is relatively less than the model developed from SFS1 data. Accuracy visualization has been presented in Figure 3. Based on Figure 3, the SFS2 model is relatively more stable than the SFS1 model. In this research, it can be seen that the average number of rules influences the resulting accuracy. These results are supported by research [37], that simple rules produce the best accuracy.

TABLE IX
AVERAGE ACCURACY AND NUMBER OF RULES

Data from model	**Accuracy**	**Number of rules**
SFS1	77.26	3
SFS2	92.01	2

IV. CONCLUSION

This study has generated two models for prioritizing the labeling of PPE distribution areas using Sort Filter Skyline (SFS). The first model (SFS1) was optimized by limiting priority areas to 50% of the total data. The second model (SFS2) further optimized SFS1 by selecting areas based on distance. SFS2 emerged as the superior model as it could choose fewer PPE distribution areas than SFS1. The rules generated by the C5.0 Algorithm can serve as knowledge for recommending priority regions for PPE distribution. The optimal model is formed from the labeling results of SFS2. The average accuracy value of SFS2 data is 92.01%, with an average of two rules. This research can be used as a model for implementing the concept of cross-regional collaboration. The implementation of cross-regional collaboration allows each region to help each other in handling disasters or pandemics. Areas that have more PPE available can help regions that have little PPE. Future research is expected to incorporate data on PPE availability in a region to assess the PPE needs in that area. Subsequent studies could optimize the SFS model by measuring the distance of a region using Google Maps technology, allowing for direct application in the PPE distribution process.

REFERENCES

[1] Q. Aini, R. R. Fauzi, and E. Khudzaeva, "Economic Impact due Covid-19 Pandemic: Sentiment Analysis on Twitter Using NaÃ¯ve Bayes Classifier and Support Vector Machine," *JOIV : International Journal on Informatics Visualization*, vol. 7, no. 3, pp. 733–741, Sep. 2023, doi: 10.30630/joiv.7.3.1474.

[2] V. Purwayoga, "Modified skyline query to measure priority region for personal protective equipment recipient of COVID-19 health workers," *Jurnal Teknologi dan Sistem Komputer*, vol. 9, no. 3, pp. 167–173, Jul. 2021, doi: 10.14710/jtsiskom.2021.14003.

[3] M. A. Shereen, S. Khan, A. Kazmi, N. Bashir, and R. Siddique, "COVID-19 infection: Emergence, transmission, and characteristics of human coronaviruses," *J Adv Res*, vol. 24, pp. 91–98, Jul. 2020, doi:10.1016/j.jare.2020.03.005.

[4] N. Novarisa, H. Helda, and R. Mulyadi, "Indonesia's COVID-19 Trend After the End of a Public Health Emergency of International Concern: Preparation for an Endemic," *Kesmas: Jurnal Kesehatan Masyarakat Nasional*, vol. 18, no. sp1, p. 25, Sep. 2023, doi:10.21109/kesmas.v18isp1.7018.

[5] Y. Liu *et al.*, "Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals," *Nature*, vol. 582, no. 7813, pp. 557–560, Jun. 2020, doi:10.1038/s41586-020-2271-3.

[6] D. Ramakrishnan, "COVID-19 and Face Masks – To Use or Not to Use!," *Indian J Community Health*, vol. 32, no. 2 (Supp), pp. 240–243, Apr. 2020, doi: 10.47203/IJCH.2020.v32i02SUPP.012.

[7] I. Tubert-Brohman, W. Sherman, M. Repasky, and T. Beuming, "Improved Docking of Polypeptides with Glide," *J Chem Inf Model*, vol. 53, no. 7, pp. 1689–1699, Jul. 2013, doi: 10.1021/ci400128m.

[8] T. M. Cook, “Personal protective equipment during the coronavirus disease (COVID) 2019 pandemic – a narrative review,” *Anaesthesia*, vol. 75, no. 7, pp. 920–927, Jul. 2020, doi: 10.1111/anae.15071.

[9] M. S. Sinha, F. T. Bourgeois, and P. K. Sorger, “Personal Protective Equipment for COVID-19: Distributed Fabrication and Additive Manufacturing,” *Am J Public Health*, vol. 110, no. 8, pp. 1162–1164, Aug. 2020, doi: 10.2105/AJPH.2020.305753.

[10] E. Furman *et al.*, “Prediction of personal protective equipment use in hospitals during COVID-19,” *Health Care Manag Sci*, vol. 24, no. 2, pp. 439–453, Jun. 2021, doi: 10.1007/s10729-021-09561-5.

[11] S. Munir and M. Asqia, “Implementasi Skyline Query pada Sistem Rekomendasi Pemilihan Tempat Kuliner di Kota Depok, Bogor, dan Tangerang,” *Jurnal Teknologi Terpadu*, vol. 7, no. 2, pp. 113–119, Dec. 2021, doi: 10.54914/jtt.v7i2.440.

[12] V. Purwayoga and B. Susanto, “Rekomendasi Daerah Penyalur Tenaga Kesehatan Covid-19 Dengan Menggunakan Skyline Query,” *Fountain of Informatics Journal*, vol. 7, no. 1, p. 22, Oct. 2021, doi:10.21111/fij.v7i1.5720.

[13] V. Purwayoga, M. Al Husaini, and H. H. Lukmana, “Visualisasi Skyline Query untuk Distribusi Tenaga Kesehatan COVID-19,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 1, Apr. 2023, doi: 10.28932/jutisi.v9i1.5624.

[14] C. B. Biddell *et al.*, “Cross-sector decision landscape in response to COVID-19: A qualitative network mapping analysis of North Carolina decision-makers,” *Front Public Health*, vol. 10, Aug. 2022, doi:10.3389/fpubh.2022.906602.

[15] I. Falagara Sigala, M. Sirenko, T. Comes, and G. Kovács, “Mitigating personal protective equipment (PPE) supply chain disruptions in pandemics – a system dynamics approach,” *International Journal of Operations & Production Management*, vol. 42, no. 13, pp. 128–154, Dec. 2022, doi: 10.1108/IJOPM-09-2021-0608.

[16] G. Baloch, F. Gzara, and S. Elhedhli, “Covid-19 PPE distribution planning with demand priorities and supply uncertainties,” *Comput Oper Res*, vol. 146, p. 105913, Oct. 2022, doi:10.1016/j.cor.2022.105913.

[17] A. T. C. Onstein, M. Ektesaby, J. Rezaei, L. A. Tavasszy, and D. A. van Damme, “Importance of factors driving firms’ decisions on spatial distribution structures,” *International Journal of Logistics Research and Applications*, vol. 23, no. 1, pp. 24–43, Jan. 2020, doi:10.1080/13675567.2019.1574729.

[18] A. Annisa and S. Khairina, “Location Selection Based on Surrounding Facilities in Google Maps using Sort Filter Skyline Algorithm,” *Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika*, vol. 7, no. 2, pp. 65–72, Jul. 2021, doi: 10.23917/khif.v7i2.12939.

[19] S. Dwiasnati and Y. Devianto, “Classification of Flood Disaster Predictions using the C5.0 and SVM Algorithms based on Flood Disaster Prone Areas,” *International Journal of Computer Trends & Technology*, vol. 67, no. 07, pp. 49–53, Jul. 2019, doi:10.14445/22312803/IJCTT-V67I7P107.

[20] I. Z. P. Hamdan, M. Othman, Y. M. Mohmad Hassim, S. Marjudi, and M. Mohd Yusof, “Customer Loyalty Prediction for Hotel Industry Using Machine Learning Approach,” *JOIV : International Journal on Informatics Visualization*, vol. 7, no. 3, pp. 695–703, Sep. 2023, doi:10.30630/joiv.7.3.1335.

[21] A. Nurkholis, Styawati, V. Purwayoga, H. H. Lukmana, A. Prihandono, and W. Koswara, “Analysis of Weather Data for Rainfall Prediction using C5.0 Decision Tree Algorithm,” in *2022 2nd International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)*, IEEE, Dec. 2022, pp. 551–555. doi:10.1109/ISMODE56940.2022.10180907.

[22] R. I. Komaraasih, I. S. Sitanggang, A. Annisa, and M. A. Agmalaro, “Sentinel-1A image classification for identification of garlic plants using decision tree and convolutional neural network,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 4, p. 1323, Dec. 2022, doi: 10.11591/ijai.v11.i4.pp1323-1332.

[23] B. R. Devi, K. Nageswara Rao, S. P. Setty, M. N. Rao, and A. Prof, “Disaster Prediction System Using IBM SPSS Data Mining Tool,” *International Journal of Engineering Trends and Technology (IJETT)*, [Online]. Available: http://www.ijettjournal.org

[24] Z. Guo, Y. Shi, F. Huang, X. Fan, and J. Huang, “Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management,” *Geoscience Frontiers*, vol. 12, no. 6, p. 101249, Nov. 2021, doi: 10.1016/j.gsf.2021.101249.

[25] R. Agramanisti Azdy and F. Darnis, “Use of Haversine Formula in Finding Distance Between Temporary Shelter and Waste End Processing Sites,” *J Phys Conf Ser*, vol. 1500, no. 1, p. 012104, Apr. 2020, doi: 10.1088/1742-6596/1500/1/012104.

[26] E. Maria, E. Budiman, Haviluddin, and M. Taruk, “Measure distance locating nearest public facilities using Haversine and Euclidean Methods,” *J Phys Conf Ser*, vol. 1450, no. 1, p. 012080, Feb. 2020, doi: 10.1088/1742-6596/1450/1/012080.

[27] Istiadi *et al.*, “Classification of Tempeh Maturity Using Decision Tree and Three Texture Features,” *JOIV : International Journal on Informatics Visualization*, vol. 6, no. 4, p. 883, Dec. 2022, doi:10.30630/joiv.6.4.983.

[28] F. Rahmad, Y. Suryanto, and K. Ramli, “Performance Comparison of Anti-Spam Technology Using Confusion Matrix Classification,” *IOP Conf Ser Mater Sci Eng*, vol. 879, no. 1, p. 012076, Jul. 2020, doi:10.1088/1757-899X/879/1/012076.

[29] J. Kludas *et al.*, “Machine Learning of Protein Interactions in Fungal Secretory Pathways,” *PLoS One*, vol. 11, no. 7, p. e0159302, Jul. 2016, doi: 10.1371/journal.pone.0159302.

[30] A. Nurkholis and I. S. Sitanggang, “A spatial analysis of soybean land suitability using spatial decision tree algorithm,” in *Sixth International Symposium on LAPAN-IPB Satellite*, T. D. Pham, K. D. Kanniah, K. Arai, G. J. P. Perez, Y. Setiawan, L. B. Prasetyo, and Y. Murayama, Eds., SPIE, Dec. 2019, p. 65. doi: 10.1117/12.2541555.

[31] N. F. Muhamad Krishnan, Z. A. Zukarnain, A. Ahmad, and M. Jamaludin, “Predicting Dengue Outbreak based on Meteorological Data Using Artificial Neural Network and Decision Tree Models,” *JOIV : International Journal on Informatics Visualization*, vol. 6, no. 3, p. 597, Sep. 2022, doi: 10.30630/joiv.6.2.788.

[32] J. Liu *et al.*, “Optimized Query Algorithms for Top- K Group Skyline,” *Wirel Commun Mob Comput*, vol. 2022, pp. 1–11, Jan. 2022, doi:10.1155/2022/3404906.

[33] A. Annisa and L. Angraeni, “Location Selection Query in Google Maps using Voronoi-based Spatial Skyline (VS2) Algorithm,” *Jurnal Online Informatika*, vol. 6, no. 1, p. 25, Jun. 2021, doi:10.15575/join.v6i1.667.

[34] F. Li, “Logistics Distribution Path Optimization Algorithm Based on Intelligent Management System,” *Comput Intell Neurosci*, vol. 2022, pp. 1–12, Sep. 2022, doi: 10.1155/2022/3699990.

[35] F. Ren, Z. Tian, J. Pan, and Y. Chiu, “Cross-regional comparative study on energy efficiency evaluation in the Yangtze River Basin of China,” *Environmental Science and Pollution Research*, vol. 27, no. 27, pp. 34037–34051, Sep. 2020, doi: 10.1007/s11356-020-09439-z.

[36] J. Li, J. Wang, H. Lee, and X. Zhao, “Cross-regional collaborative governance in the process of pollution industry transfer: The case of enclave parks in China,” *J Environ Manage*, vol. 330, p. 117113, Mar. 2023, doi: 10.1016/j.jenvman.2022.117113.

[37] J. L. Delgado-Gallegos *et al.*, “Application of C5.0 Algorithm for the Assessment of Perceived Stress in Healthcare Professionals Attending COVID-19,” *Brain Sci*, vol. 13, no. 3, p. 513, Mar. 2023, doi:10.3390/brainsci13030513.