Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 3 (2023) 733 - 742 ISSN Media Electronic: 2580-0760

Clustering Analysis and Mapping of ISPA Disease Spread Patterns in Bireuen District

Mutammimul Ula[1], Tsania Asha Fadilah Daulay[2], Richki Hardi[3], Sujacka Retno[4], Angga Pratama[5], Ilham Sahputra[6]
[1,2,5,6]Sistem Informasi, Teknik, Universitas Malikussaleh
[3]Teknik Informasi, Universitas Mulia
[4]Teknik Informatika, Teknik, Universitas Malikussaleh
[1]mutammimul@unimal.ac.id, [2]niashafadilah26@gmail.com, [3] richki@universitasmulia.ac.id, [4]sujacka@unimal.ac.id, [5]anggapratama@unimal.ac.id, [6]lham.sahputra@unimal.ac.id

Abstract

ISPA disease can be detected by analyzing the regional distribution map of the disease. Early detection of ARI is very important for effective prevention. The study conducted in Bireuen Regency used data from 2019 to 2021, sourced from dr. Fauziah Bireuen Hospital, revealed that there was an increase in ARI cases from an average of 13.18 to 59.24 per year. The aims of the study were to identify ARI clusters, analyze disease patterns using Spatial Pattern Analysis and Flexible Shaped Spatial Scanning Statistics. The methodology involves collecting patient data for each ARI case and processing it using DBSCAN to obtain cluster points on the map. Spatial Pattern Analysis is used to analyze these clusters and identify hotspot points on the map. The analysis resulted in four clusters: Cluster 1 (6 subdistrict), Cluster 2 (4 subdistrict), Cluster 3 (1 subdistrict), and Cluster 4 (6 subdistrict). The study identified 6 hotspots in 2019, 5 hotspots in 2020, and 6 hotspots in 2021. Each ARI disease clustering map shows the distribution of ARI cases and identifies areas prone to the disease. These findings provide valuable insights for targeted interventions and preventive actions in identified high-risk areas of ISPA.

Keywords: *clustering; DBCAN; distribution pattern detection; ISPA; mapping*

1. Introduction

ISPA is an infectious disease characterized by acute infection of the respiratory [1]. One of the factors that causes the transmission of ISPA is the density of residence [2]. In addition, according to WHO, the lack of infection control in the distribution is also becomes a factor that cannot be underestimated by the relevant agencies [3], [4].

Machine learning is an extraction process algorithm [5],[6]. The process of extracting data to find information from existing datasets using clustering method to detect the patterns of disease distribution [7],[8]. The DBSCAN algorithm is used as a clustering analysis calculation. Meanwhile, Spatial Pattern Analysis and Flexibly Shaped Spatial Scan Statistics are used to detect the distribution patterns [9]–[11].

Clustering is a form of analysis of object density in finding patterns of disease distribution in sub-district points to determine and analyze the distance measure of one point to another by using the clustering method [12]–[16]. A model that can see the number of ISPA patients and the factors that are hypothesized has an influence to it by combining the Flexibly Shaped Spatial Scan Statistics model and mapping. This model has a process of mean and variance values of the response variables for the detection patterns of the disease distribution [17]–[21].

The research of Teguh et.al., There were five clusters of ISPA disease in 2020, according to a study of illness distribution data using a density-based spatial clustering technique with noise. The highest areas found in Cluster 1 are Purwakarta and Babakancikao [22],[23]. In the research conducted by Niswatul, et al, carried out the detection analysis of a pattern of ISPA distribution as an Impact of the Oil and Gas Industry by using Spatial Pattern Analysis and Flexibly Shaped Spatial Scan Statistics, there are sub-districts that are most affected by ISPA and these areas are included in quadrant I (HH), namely Trucuk, Bojonegoro and Kapas.

The test of Moran's I index resulted in spatial autocorrelation in 2012 and 2013 has a significant value at $\alpha > 5\%$, in 2014 and 2015 also has a significant at $\alpha >$

Accepted: 14-02-2023 | Received in revised: 17-05-2023 | Published: 05-06-2023

5%, no more than 15%. The sub-districts that become hotspots for ISPA disease include areas that are included in quadrant I (HH), namely Trucuk, Bojonegoro and Kapas. By using the Flexible Shaped Spatial Scan Statistical method, it obtained 3 significant ISPA criterias [23].

The next research in Modeling and Mapping the number of DBD (Dengue Fever) Cases in Surabaya City using the Geographicaly Binomial Regression model and Flexibly Shaped Spatial Scan Statistics. The results of this research is to find the areas affected by DDB with three clusters namely vulnerable, moderate and safe areas. The sub-district that the most affected by DHF is the Benowo area which has a relative risk value of 2.34. Based on the results of the elaboration of previous research, clustering analysis can be carried out in mapping by using the DBSCAN method and for the detection of distribution patterns is using a combination of the two methods, namely Spatial Pattern Analysis and Flexibly Shaped Spatial Scan Statistics.

This research aimed to find out the clusters and distribution patterns in the district of Bireuen. The results of this study are expected to be used as a reference in policies on preventing the spread of ARI. The data used in this study were ISPA patient data taken from the medical records of dr. Fauziah Hospital, Bireuen District, Aceh.

2. Research Methods

This research was conducted at dr. Fauziah Hospital, Bireuen. The data will be collected, processed and analyzed to obtain results in a form of information that will be used for mapping the ISPA diseases in Bireuen Regency. The presentation of this data will be formed in an image visualization of mapping the ISPA disease.

The data used in this research were ISPA patient data taken from the medical records of dr. Fauziah Hospital, Bireuen Regency, Aceh. The data will be analyzed using the three methods that have been described.

2.1. Research Framework of DBSCAN

Data analysis was carried out in three steps. The first step is mapping. At this step, mapping is carried out by clustering the data by using DBSCAN to obtain the mapping results for each region or area where the research is conducted. The next step is carried out by analyzing the clustering data by using the Density Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm. By calculating the eps, minpts and neighborhood to determine homogeneous clusters. After that mapping can be carried out from the results of cluster analysis. The framework research of DBSCAN can be seen in Figure 1.

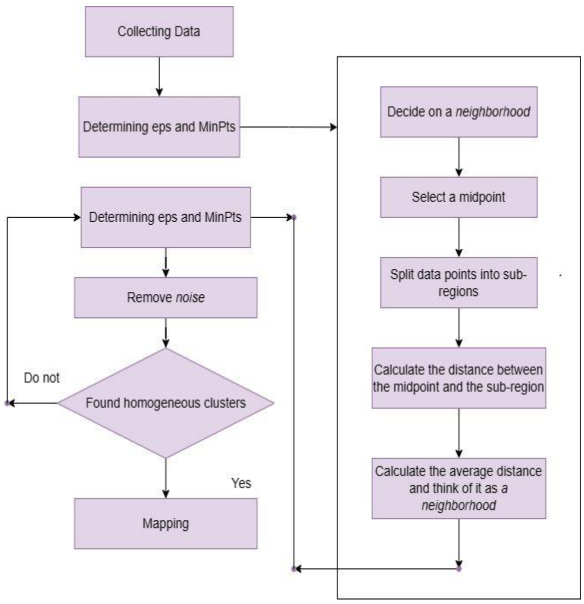

Figure 1. ISPA Disease DBSCAN Mapping Steps

Based on Figure 1, the first thing to do in the research was to collect the ISPA data at dr. Fauziah. Hospital Then the data obtained is formed into a dataset so that it can be analyzed using Python. Then determine the eps, which is the desired density and minpts, which is the minimum number of samples for each calculated data.

The method for determining eps and minpts can go through steps including by deciding on a neighborhood, then choosing a centroid, dividing the data points into sub-regions, then calculating the distance between the centroid and the sub-region, then calculating the distance and average to find a neighborhood.

After finding the neighborhood, then use the Density Based Spatial Clustering of Application with Noise method to the data. Removing noise and finding homogeneous clusters or the same group will form clusters. If no clusters formed, then repeat the DBSCAN clustering steps again and if it's has a cluster formed, you can immediately do the mapping.

The second step is the diagnosis of distribution patterns by using the Spatial Pattern Analysis and Flexibly Shaped Spatial Scan Statistics methods. This step is carried out to find out the diagnosis in distribution of patients. So it will provide the pattern of the spread of ISPA in the area researched.

The step begins with measuring the similiarity of the symptoms of the disease. Furthermore, measurements of local similarity and global similarity were carried out to find the highest similarity of the data. All stages of this diagnosis are stored in the base case.

The research scheme at this step can be seen in Figure 2.

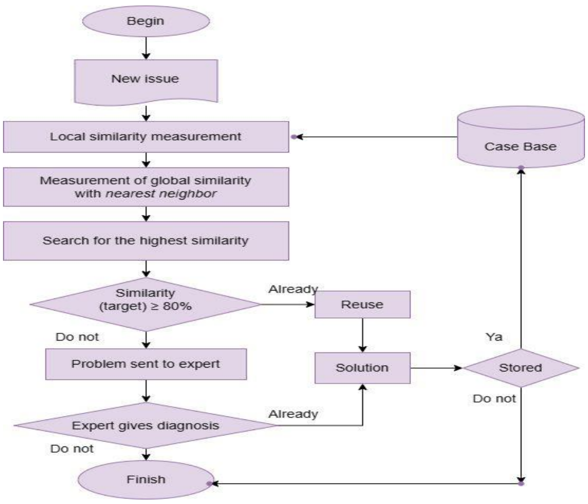

Figure 2. Diagnosis Step

In Figure 2, there are the steps for carrying out the diagnosis. Starts with finding a new problem, namely the ISPA disease, then measuring the local similarity to find the highest similarity data. The problems found are sent out to the experts to find a diagnosis of it. The results are stored in a base case where in this case is a database to store the data. ISPA disease data obtained from the diagnosis is used for the final step.

The last step is to combine the first and second step so that a more complex data analysis scheme is obtained. This step is will be used as a reference for the researchers in conducting research to find a maximum results. This combined scheme can be seen in Figure 3.

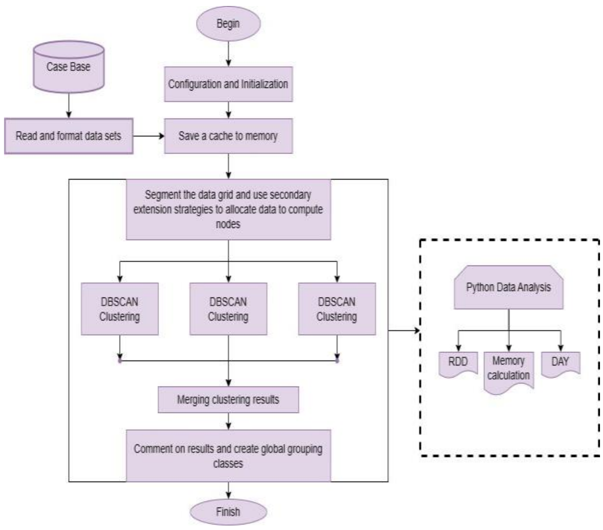

Figure 3. Mapping and Diagnostics Combined

In Figure 3 explained that at this step the clustering results from database and the diagnoses in the base case are processed by using Python programming. The results of data collection and the analysis of the results by using Python will produce a visualization of information in form of a map.

At this step, the calculation will be carried out to process and analyze the data so it can be used to obtain the information needed. Through this step, the data will be processed in three methods according to the research framework. The results of the data analyzed will provide an information about the ISPA disease mapping in Bireuen Regency.

2.2. Calculation Formulas

DBSCAN is a clustering algorithm that carried out with datasets based on the density of distances between objects [24].

The advantage of DBSCAN is it can detect outliers/noises. Formula 1 is the calculation of the distances on DBSCAN is using the Euclidean Distance formula.

$$Distance = \sqrt{(x - xp)^2 + (y - yp)^2} \quad (1)$$

with x and y as coordinates of the axes of the destination point. Xp and yp as the coordinates of the axis point.

Spatial Pattern is the result of a physical or social process of any location that occurs on the surface of the earth. The statistical concept of a pattern shows how geographical objects are distributed [25].

The calculation of the value of Moran's I is stated in Formula 2:

$$I = \frac{N}{\sum i \sum j\, wij} \frac{\sum i \sum j\, wij(Xi - \bar{X})(Xi - \bar{X})}{\sum i\, (Xi - \bar{X})^2} \quad (2)$$

with Moran's Value denoted by I. N represents the number of incident locations. Values at the event location are denoted by xj and xi. $\bar{X}$ is the average number of all variables. And wij is a weighting element between the incident locations.
Then carry out the calculations for the statistical expectation value E(I) in Formula 3.

$$E(I) = \frac{1}{(n-1)} \quad (3)$$

where E(I) is the statistical expectation value. And N is the number of the events. Next, calculate the value of the continuity element C in Formula 4.

$$C = \sum_{i=1}^{n} C \sum_{j=1}^{n} Cij$$

$$S1 = \frac{\sum_{i=1}^{n} C \sum_{j=1}^{n} (Cij + Cji)^2}{2} \quad (4)$$

where Cij is the value of the continuity element. While the total value of the incident location is denoted by Ci and Cj. Then do the calculation for the variance value Var(I) in Formula 5.

$$Var(I) = \frac{n^2 S - nS^2 + 3(C)^2}{(C)^2 (n^2 - 1)} \quad (5)$$

DOI: https://doi.org/10.29207/resti.v7i3.4936

where Var(I) is the variance value. The number of the events are denoted by n. S represents the standard deviation and C is the value of the continuity element. Then perform the statistical tests on the parameters I_i based on the conditions if H_0 is I_i equal to $E(I_i)$ then there is no spatial autocorrelation at location *-i*. If H_1 is I_i not equal to $E(I_i)$ then there is spatial autocorrelation at location *-i*.

To find out the distribution pattern formed based on the index value can be seen in table 1 [26].

Table 1. Moran's I Index

Value I	Information
I > 0	Cluster Patterns (has many values in common)
I < 0	Random Patterns (value is unclear)
I = 0	Spread Patterns (high and low values spread over the data)

Then the next equation obtained can be seen in formula 6.

$$Z(I) = \frac{I - E(I)}{\sqrt{Var\ (I)}} \tag{6}$$

where I is Moran's value. Z(I) is the value of the statistic test. E(I) is the expected value of the test and Var(I) is the variance.

Flexibly Shaped Spatial Scan Statistics is a statistical technique for detecting clusters in form of aggregated data. The statistical test used the hypothesis of Monte Carlo approach with the condition if H_0 is $E(X(z))$ equal to $\mu(z)$ for all Z where area Z is not an area of the distribution of ISPA. If H_1 is $E(X(z))$ higher than $\mu(z)$ for some Z where area Z is an area of the distribution of ISPA disease.

For each Z, the likelihood was calculated to observe the number of hotspots inside and outside Z areas. Statistical tests were conducted by using likelihood ratio tests as seen in Formula 7.

$$\lambda = \left[\frac{x(z)}{\mu(z)}\right]^{x(z)} \left[\frac{x(Z^c)}{\mu(Z^c)}\right]^{x(z^c)} I\left[\frac{x(z)}{\mu(z)} > \frac{x(z^c)}{\mu(z^c)}\right] \tag{7}$$

where λ which represents the statistic tests. Z^c represents the area outside the Z area. The total average number of spreading points is denoted by x(.) and the indicator function is denoted by I(.) with 1 value if Z has a greater probability and 0 otherwise.

To find the predicted value or p-value, the first step is to find the lowest log likelihood ratio value t_0. The second step is to create a random data of the same size on the real dataset under conditions H_0. The third step is to find a random data values by scanning the Z window based on conditions H_0. The fourth step is to take the log likelihood ratio value from the scanning window and then calculate the highest log likelihood ratio value from the first random data simulation. The next step is repeatoing the second to fourth steps m times of iteration or simulation, to obtain the highest number of log likelihood ratio values from random data and the real data.

Next, calculations are carried out to calculate the p-value can be seen in Formula 8.

$$p - value = \frac{N\ (T_{(x)} \geq t_0)}{m+1} \tag{8}$$

where T represents the sum of the log values. Data values are denoted by x. t_0 is the log likelihood value that has the highest ratio. And m is the number of simulations carried out in the research.

2.3. Dataset

The dataset used in this research is the data that has been collected from the medical records of ISPA patients at dr. Fauziah Hospital in Bireuen. The data collected consists of 3 years, namely 2019, 2020 and 2021. The dataset for DBSCAN calculations uses patient address data. Where sub-district data is converted into a latitude data and longitude so that the clustering calculations can be carried out. The dataset for clustering calculations can be seen in table 2.

Table 2. ISPA location Dataset

No	Sub-District	latitude	longitude
P-1	Kuala	5.2165785	96.7197807
P-2	Peudada	5.0718905	96.5684513
P-3	Peusangan	5.1975061	96.7781702
P-4	Juli	5.0807985	96.6990411
P-5	Gandapura	5.2264184	96.8818435
P-6	Kota Juang	5.1958923	96.7203622
P-7	Jangka	5.2281504	96.7954511
…	…	…	…
…	…	…	…
P-147	Jeumpa	5.1046072	96.6389621

The dataset used in the calculations for the detection of distribution patterns. Where the data used is the number of cases, the dataset used is the number of cases of ISPA in each sub-district. The dataset used for Moran's I calculations and the Monte Carlo statistical test is the dataset of the number of cases per district.

The dataset used for calculating SHP in mapping. Where the keywords in the dataset must be adjusted to the keywords in the data in Bireuen map table. The dataset included in the SHP map can be seen in table 3.

Table 3. ISPA Dataset

FID	Sub-District	ISPA 2019	ISPA 2020	ISPA 2021	ISPA 2022
19	Peusangan Selatan	7	8	26	41
20	Peusangan siblah krueng	9	12	42	63
21	Juli	29	20	104	153
22	Peulimbang	11	8	38	57
23	Pandrah	14	17	110	141
24	Jeunieb	52	46	235	333
26	Peudada	26	18	70	114
28	Kotajuang	13	5	40	58
29	Makmur	7	5	13	25
32	Jeumpa	14	13	42	69

DOI: https://doi.org/10.29207/resti.v7i3.4936

FID	Sub-District	ISPA 2019	ISPA 2020	ISPA 2021	ISPA 2022
33	Simpang mamplam	4	3	15	22
36	Peusangan	36	42	155	233
38	Kuala	2	4	14	20
42	Kutablang	2	2	12	16
44	Gandapura	8	10	33	51
45	Jangka	11	3	24	38

The dataset in table 3 is ready to be processed and analyzed to obtain the calculation results and to find a useful information which is the result of research. Datasets processed by using Python are converted into csv format and datasets for SHP used an excel-97 format. For pattern detection calculations, you can use the conventioanl Excel format.

3. Results and Discussion

Mapping systems or known as Geographic Information Systems are widely used in health sector. For example, on the map of disease distribution patterns. Analysis in mapping system is carried out by processing data from various regions.

The data collected is in form of the number of cases of ISPA disease in a certain area. Based on the data obtained, a mapping system can be created so that information on the data can be used for decision making and policy.

There are many clustering methods that are often used for the clustering process to map the distribution of a disease. One of the methods used is Density Based Spatial Clustering of Application with Noise or known as DBSCAN. In this research, researcher used DBSCAN, the results will be obtained in form of grouping based on clusters in the affected areas. Then the number of clusters with the distribution area of the disease is obtained. However, it is not possible to know which indications are safe or unsafe in the distribution area.

To find out the risk of distribution and prone of a disease area, two methods are used together, namely Spatial Pattern Analysis and Flexibly Shaped Spatial Scan Statistics. By combining these two methods, spatial autocorrelation and hotspot detection can be carried out to determine the indications of areas that are safe, prone and unsafe for a disease as well as the relative amount of risk in each type of these areas. However, it is not possible to know the distribution cluster clusters in ISPA disease.

3.1. Data Analysis and Calculations

DBSCAN calculations used the data on the number of ISPA cases in dr. Fauziah Hospital, Bireuen over of 3 years period. The 3 years taken are from 2019 to 2021. The data of ISPA disease is obtained from patient medical records and processed by using the DBSCAN method to obtain the clustering results.

The variable used to determine clustering on the mapping is the district area. There are a total of 17 sub-districts in Bireuen Regency. These sub-districts are Gandapura, Jangka, Jeumpa, Jeunieb, Juli, Kota Juang, Kuala, Kuta Blang, Makmur, Pandrah, Peudada, Peulimbang, Peusangan, Peusangan Selatan, Peusangan Siblah Krueng, Samalanga, and Simpang Mamplam.

Table 4. Initial Centroid

Initial Cluster	Centroid
C1	Gandapura
C2	Jangka
C3	Jeumpa
C4	Jeunieb
C5	Juli
C6	Kota Juang
C7	Kuala
C8	Kuta Blang
C9	Makmur
C10	Pandrah
C11	Peudada
C12	Peulimbang
C13	Peusangan
C14	Peusangan Selatan
C15	Peusangan Siblah Krueng
C16	Samalanga
C17	Simpang Mamplam

Based on Table 4, it obtains a number of clusters formed based on the minimum number of samples and epsilon that has been determined, namely into 4 groups divided into 3 clusters and 1 noise. In addition to the number of clusters that are formed, the results of Density Based Spatial Clustering of Application with Noise calculations also show which points are included in each cluster. The results of the clusters and the centroid formed can be seen in table 5.

Table 5. Clustering Results

Grouping Cluster	Centroid
Cluster 0 (noise)	C1, C2, C6, C7, C8, C13
Cluster 1	C4, C10, C11, C12,C16, C17
Cluster 2	C3, C5, C14, C15
Cluster 3	C9

In this method, the calculation is continued by calculating the statistical data per year. So in this method and also the next method, the data is processed based on each year, namely ISPA disease data for 2019, ISPA disease data for 2020, and ISPA disease data for 2021 in all sub-districts in Bireuen Regency.

In 2019, there were 4 sub-districts in Bireuen which had a very high number of cases, namely Kota Juang sub-district with 49 cases. Then Peusangan sub-district with a total of 33 cases. Jeumpa sub-district with 27 cases and Kuala sub-district with 26 cases.

Statistical data on ISPA cases from 2019 to 2021 in Bireuen Regency can be seen in table 6.

Based on table 6, it shows that ISPA cases in Bireuen experienced a slight decrease from 2019 to 2020 with an average number of cases from 13.47 to 13.18.

However, ISPA cases experienced a very high increase in 2021 with an average of 59.24. Where in 2021 there are 8 sub-districts with a very high number of cases, namely Kota Juang sub-district with 235 cases. Next is Peusangan sub-district with 181 cases. Then Juli sub-district with 110 cases. Jeumpa sub-district with 104 cases. Then Kuala subdistrict with 70 cases. Furthermore, Jangka and Peudada, each of it has 42 cases.

Table 6. Statitical data on ISPA cases

Year	Average	Median	Std Deviation	Min	Max
2019	13,47	9,00	12,99575723	2,00	49,00
2020	13,18	8,00	13,95186051	2,00	48,00
2021	59,24	38,00	64,15949794	8,00	235,00

Next, the calculation is carried out by using the Moran I index. The results of the data analyzed by using the Moran's I Index can be seen in table 7.

Table 7. Moran I index Results

Year	I	E(I)	var(I)	Z_{value}	P_{value}
2019	-0,2334	-0,625	0,0325	-0,94795	0,34315
2020	-0,3224	-0,625	0,0322	-1,44807	0,14759
2021	-0,1665	-0,625	0,0303	-0,55061	0,55061

Next, the spatial autocorrelation process is carried out. Where spatial autocorrelation is a measure or correlation between an object variable and itself based on distance, time and region. The results of spatial autocorrelation process on ISPA disease data in Bireuen can be seen in table 8.

Table 8. Spatial Autocorrelation Results

Year	H_0 / H_1	Information
2019	H_1	Has a spatial autocorrelation
2020	H_1	Has a spatial autocorrelation
2021	H_1	Has a spatial autocorrelation

From the results of table 8, it can be seen that the data for 2019, 2020 and 2021 have the same information, namely there is a spatial autocorrelation. So the next step of data processing can be carried out. The Moran's I index value is in the range of -1 to 1, it is known that the spatial dependency between sub-districts related to ISPA cases is low. The results of the significance testing step to determine the distribution pattern can be seen in table 9.

Table 9. Distribution Pattern of ISPA disease

Year	I	I_0	Correlation I with I_0	Conclusion
2019	-0,233	-0,625	$I > I_0$	Cluster Pattern
2020	-0,322	-0,625	$I > I_0$	Cluster Pattern
2021	-0,166	-0,625	$I > I_0$	Cluster Pattern

From table 9, it can be seen the pattern of the distribution of ISPA in sub-district in Bireuen Regency in 2019, 2020 and 2021 is a cluster pattern. The clustered pattern is a pattern where the feature values in each district have a lot of similarities between each other.

The significance level of ISPA disease in Bireuen Regency was carried out by using a Monte Carlo relative risk simulation method. The simulation method was carried out with a total of 99, 999, and 9999 repetitions. The following are the results of the detection of ISPA disease with the Flexibly Shaped Spatial Scan Statistics Monte Carlo model.

Table 10. Monte Carlo Calculation Results

Year	ISPA bags	Number of Case	Relative Risk	Z_{value}	P_{value}
2019	1	248	2,7	0,604	0,107
2020	1	218	3,1	0,415	0,055
2021	1	981	3,85	0,180	0,004

In table 10, it can be seen that the number of ISPA bags produced in Bireuen Regency is 1 per year. The relative risk obtained in 2019 is 2.7. The relative risk for 2020 is 3.1. And the relative risk in 2021 is 3.85

3.2. Map Implementation Results

System implementation is a calculation of data by using analytical techniques. In DBSCAN method, the data calculation step uses Python analysis is carried out by using the Google colaboratory web-based application. Python will provides the number of clusters formed in the analysis calculation in form of a Bireuen Regency map. The results of cluster mapping by using the DBSCAN method can be seen in Figure 4.

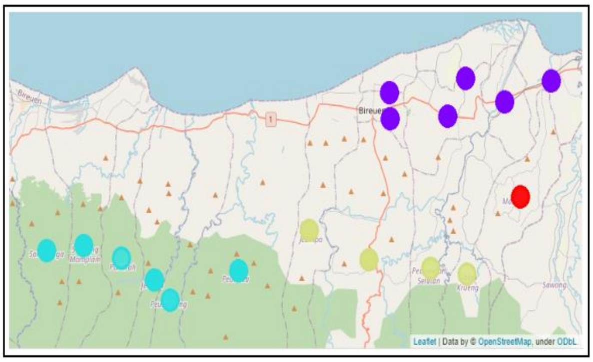

Figure 4. Bireuen Regency Cluster Mapping Result

The results in Figure 4 show that there are 4 clusters represented by 4 different colors. Cluster 0 in purple color has 6 sub-districts, namely Kuala sub-district, Kota Juang sub-district, Peusangan sub-district, Jangka sub-district, Kuta Blang sub-district and Gandapura sub-district. Cluster 1 in light blue color includes Samalanga sub-district, Simpang Mamplam sub-district, Pandrah sub-district, Jeunieb sub-district, and Peudada sub-district.

Cluster 2 in green color consists of Jeumpa sub-district, Juli sub-district, Peusangan Selatan sub-district, and Peusangan Siblah Krueng sub-district. Last, Cluster 3 consist of Makmur sub-district which is represented by the red color on the map.

DOI: https://doi.org/10.29207/resti.v7i3.4936

The Spatial Pattern Analysis method uses the ArcMap application in calculating and analyzing ISPA disease data. The results of the map that has been analyzed using Moran's Scatterplot Index can be seen in Figure 5.

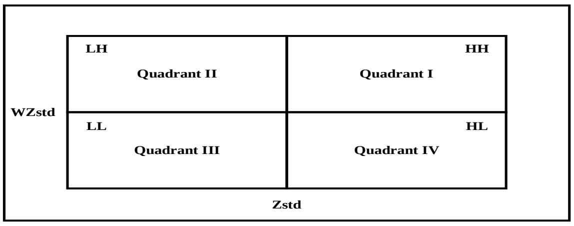

Figure 5. Moran’s Scatterplot Index

Quadrant I or also called HH (High-High) shows the areas that have high observation values and surrounded by the areas with high values too. Quadrant II is called LH (Low-High) meaning that this quadrant shows areas that have low observation values but are surrounded by areas with high observation values.

Quadrant III, also called LL (Low-Low), is a quadrant of an area with low observation values and also surrounded by areas with low observation values. Quadrant IV or called as HL (High-Low) shows areas with high observation values but surrounded by areas with low observation values.

Based on Moran's Scatterplot Index, ISPA per sub-district is distributed in Quadrant I to Quadrant IV. Quadrants I and Quadrant IV are classified as Hotspots while Quadrants II and Quadrant III are classified as Coldspots.

The results of the observations in sub-districts in Bireuen Regency on the distribution of ISPA in 2019 can be seen in Figure 6.

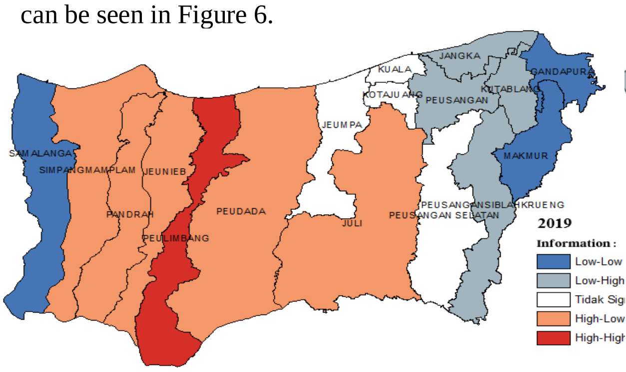

Figure 6. ISPA Disease Analysis Results in 2019

The results in Figure 6 show Quadrant I (HH) which is represented in dark red consist of Peulimbang sub-district. Quadrant II (LH) in light blue consist of Jangka, Peusangan, Kuta Blang, and Pesangan Siblah Krueng sub-district. Quadrant III (LL) in dark blue includes Samalanga, Makmur and Gandapura districts. Quadrant IV (HL) with a light orange color consists of Simpang Mamplam, Pandrah, Jeunieb, Peudada, and Juli sub-districts.

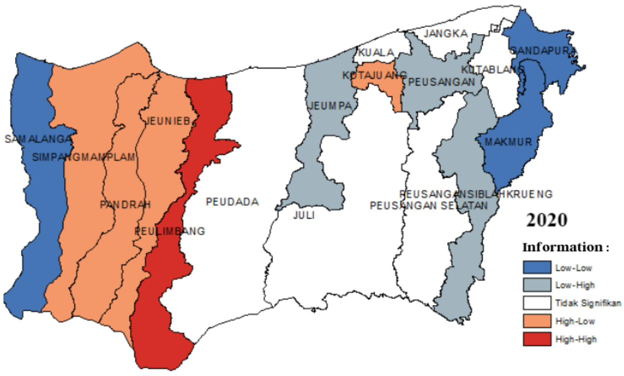

Figure 7. ISPA Disease Analysis Results in 2020

The results in Figure 7 show that Quadrant I (HH) which is represented in dark red, consist of Peulimbang sub-district. Quadrant II (LH) in light blue consist of Jeumpa, Peusangan, and Peusangan Siblah Krueng sub-districts. Quadrant III (LL), which is represented in dark blue, includes Samalanga, Makmur and Gandapura sub-districts. Quadrant IV (HL) with a light orange color consists of Simpang Mamplam, Pandrah, Jeunieb, and Kota Juang sub-districts.

The results of the observations in sub-districts in Bireuen Regency on the distribution of ISPA in 2021 can be seen in Figure 8.

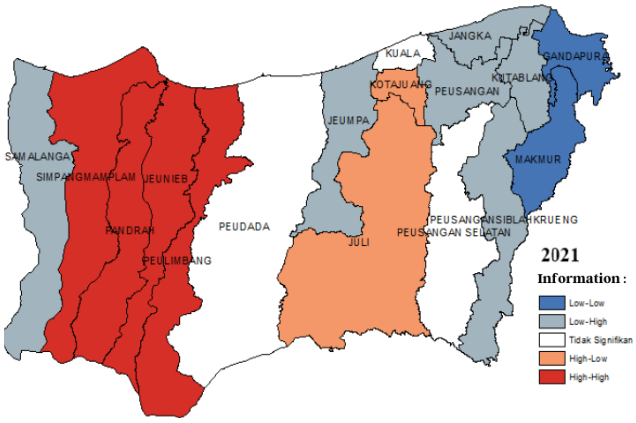

Figure 8. ISPA Disease Analysis Results in 2021

The results in Figure 8 show Quadrant I (HH) in dark red consist of Simpang Mamplam, Pandrah, Jeunieb and Peulimbang sub-districts. The light blue quadrant II (LH) covers the areas of Samalanga, Jeumpa, Peusangan, Jangka, Kuta Blang, and Peusangan Siblah Krueng sub-districts. Quadrant III (LL) in dark blue includes the districts of Makmur and Gandapura. Quadrant IV (HL), which is represented in light orange, consists of the Juli sub-district and Kota Juang.

The Flexibly Shaped Spatial Scan Statistics method also uses a mapping application called ArcMap. This method detects the areas that are hotspots for ISPA in Bireuen Regency. The territorial detection will generate a region partition. The regional partition formed will later be used as a reference in determining the prone category for areas with ISPA.

Based on the calculation of standard deviation, the category of disease susceptibility level is divided into 3 category, namely prone, moderate and safe. Prone is an area that is easily affected by disease, moderate is an area that is not prone but also not a safe area, while the safe category is an area that is not prone to ISPA disease.

The distribution of ISPA disease in 2019 in Bireuen Regency can be seen in Figure 9.

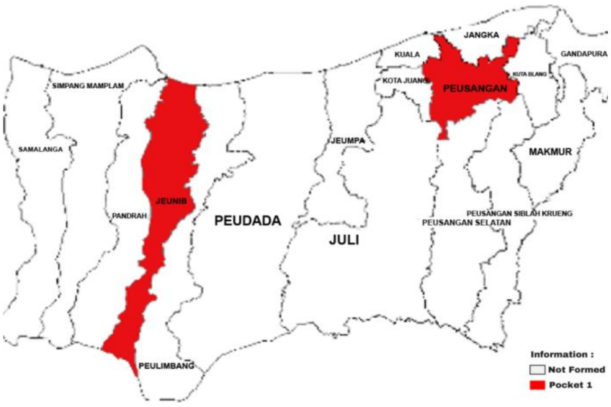

Figure 8. The distribution of ISPA disease in 2019

Based on Figure 9, it is known that there was 2 area detected where the area consist of Jeunieb sub-district and Peusangan sub-district. Furthermore, the distribution of ISPA disease in 2020 in Bireuen Regency can be seen in Figure 10.

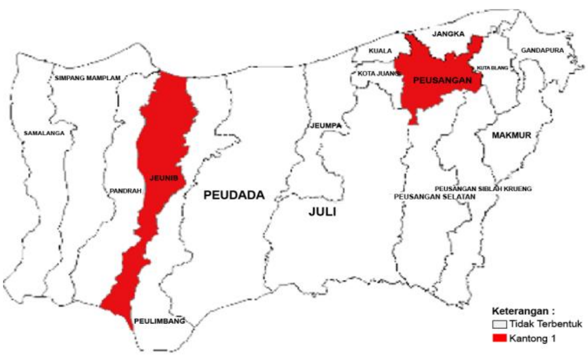

Figure 9. The distribution of ISPA disease in 2020

The results in Figure 10 show that the distribution of ISPA disease in 2020 has not changed from the distribution of ISPA disease in Bireuen Regency in the previous year, namely 2019. The distribution of ISPA disease in 2021 in Bireuen Regency can be seen in Figure 11.

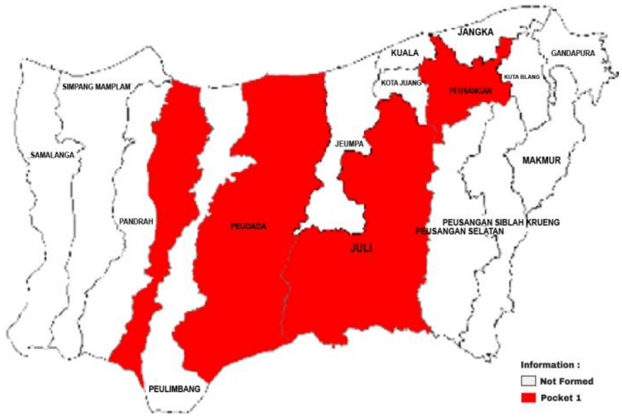

Figure 10. The distribution of ISPA disease in 2021

The results in Figure 11 show that there are 4 sub-districts included in the distribution of ISPA disease in Bireuen Regency, including Jeunieb, Peudada, Juli, and Peusangan sub-districts. From the overall results of the distribution of ISPA disease in 2019, 2020 and 2021, conclusions can be drawn about which areas are classified as prone, moderate or safe areas. The results can be seen in Figure 12.

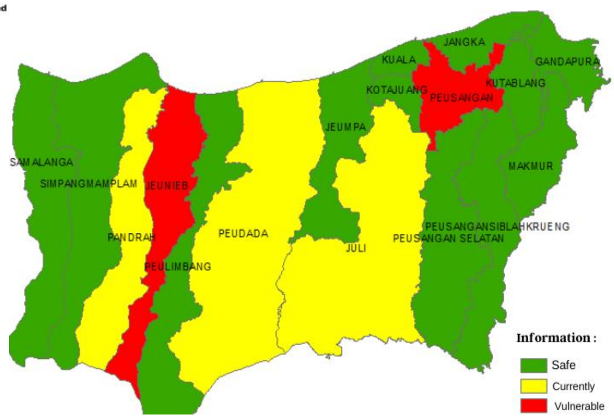

Figure 11. ISPA Disease Prone Areas based on data in 2019-2021

Based on Figure 12, it is known that the sub-districts that are classified as safe which are represented in green color on the map consist of Samalanga, Simpang Mamplam, Peulimbang, Jeumpa, Kuala, Kota Juang, Jangka, Kuta Blang, Peusangan Selatan, Peusangan Siblah Krueng, Makmur and Gandapura sub-districts. The sub-districts classified as moderate represented in colored yellow include the sub-districts of Pandrah, Peudada and Juli. Meanwhile the areas that need attention for and classified as a prone areas represented in red color consist of Jeunieb and Peusangan sub-districts.

4. Conclusion

The study's results in the distribution of the pattern of the incidence of ARI for quadrant I (HH) consisted of the mamplam, pandrah, jeunib, peulimbang intersections. Quadrant II (LH) sub-districts of

DOI: https://doi.org/10.29207/resti.v7i3.4936

Samalanga, Jeumpa, Peusangan, Jangka, Kuta Blang, and Peusangan Siblah Krueng. Quadrant III (LL) Makmur District, Gandapura. Finally, Quadrant IV (HL) July, Juang City. Research results for the last three years with 1447 dataset cases. The effects of cluster 1 with the DBSCAN model include six sub-districts, cluster 2 consists of 4 sub-districts, cluster 3 only has one sub-district and cluster 4 consists of 6 sub-districts.

Based on the discussion, suggestions that be made for further related research are observations for clustering analysis focused on coordinate data of villages in Bireuen Regency so it will produce more detailed clustering points. It is also advisable to input the population data as a comparison with the number of cases of ISPA patients.

Acknowledgment

Thanks to Universitas Malikussaleh as a place to gain knowledge and information given to the researcher to finish the research and to dr. Fauziah Hospital, Bireuen, because without the ISPA disease data given from the Hospital, this research can not be carried out.

References

[1] Ririnisahawaitun and D. A. Ikhwani, “Artikel Penelitian Pengaruh Pendidikan Kesehatan Terhadap Pengetahuan Pencegahan Penyakit Menular Ispa Pada Balita Di Desa Kalibambang Wilayah Kerja Puskesmas Lenek,” vol. 7, no. 2, 2020.

[2] N. Qona’ah, D. Sutikno, and J. Statistika, “Analisis Pola Persebaran ISPA (Infeksi Saluran Pernafasan Akut) Sebagai Dampak Industri Migas di Kabupaten Bojonegoro Menggunakan Spatial Pattern Analysis dan Flexibly Shaped Spatial Scan Statistic,” 2016.

[3] Ernyasih, F. Fajrini, and N. Latifah, “Analisis Hubungan Iklim (CurahHujan, Kelembaban, Suhu Udara dan Kecepatan Angin) dengan Kasus ISPA di DKI Jakarta Tahun 2011 – 2015,” 2018.

[4] C. Herawati and H. Sriwaty, “Analisis Perilaku Merokok, Penggunaan Anti Nyamuk Bakar Dan Penggunaan Bahan Bakar Memasak Dengan Kejadian Ispa Pada Balita,” *Jurnal Kesehatan*, vol. Vol. 9 No.1, 2018.

[5] M. Jupri and R. Sarno, “Data mining, fuzzy AHP and TOPSIS for optimizing taxpayer supervision,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 1, pp. 75–87, 2019.
https:// doi: 10.11591/ijeecs.v18.i1.pp75-87.

[6] I. W. Gamadarenda and I. Waspada, “Implementasi Data Mining Untuk Deteksi Penyakit Ginjal Kronis (Pgk) Menggunakan K-Nearest Neighbor (Knn) Dengan Backward Elimination,” vol. 7, no. 2, pp. 417–426, 2020, doi: 10.25126/jtiik.202071896.

[7] S. Ucha Putri, E. Irawan, F. Rizky, S. Tunas Bangsa, P. A. - Indonesia Jln Sudirman Blok No, and S. Utara, “Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5,” 2021.

[8] B. Tri, R. Doni, S. Susanti, and A. Mubarok, “Penerapan Data Mininguntuk Klasifikasi Penyakit Hepatocellular Carcinomamenggunakan Algoritma Naïve Bayes,” *JURNAL RESPONSIF*, vol. 3, no. 1, pp. 12–19, 2021, [Online]. Available: http://ejurnal.ars.ac.id/index.php/jti

[9] K. Mumtaz and K. Duraiswamy, “An Analysis on Density Based Clustering of Multi Dimensional Spatial Data,” *Indian Journal of Computer Science and Engineering*, vol. Vol 1 No 1, pp. 8–12, 2010, Accessed: Mar. 27, 2022. [Online].
https://www.researchgate.net/publication/45681614_An_Analysis_on_Density_Based_Clustering_of_Multi_Dimensional_Spatial_Data

[10] A. J. Curtis and W. A. A. Lee, “Spatial patterns of diabetes related health problems for vulnerable populations in Los Angeles,” *Int J Health Geogr*, vol. 9, Aug. 2010, doi: 10.1186/1476-072X-9-43.

[11] T. Tango and K. Takahashi, “A flexibly shaped spatial scan statistic for detecting clusters,” *Int J Health Geogr*, vol. 4, no. 1, pp. 1–15, May 2005,
https://doi: 10.1186/1476-072X-4-11/TABLES/6.

[12] F. Zubedi, “Analisis Cluster Hasil Try Out Siswa Mts Alhuda Gorontalo Dengan Chi-Sim Cosimilarity Dan K-Means Clustering,” 2020.

[13] F. Fania, M. Azzahra, and A. P. Windarto, “Analisis Metode Clustering Pada Wilayah BerdasarkanJenis Pencemaran Lingkungan Hidup,” 2020. [Online]. Available: https://www.bps.go.id.

[14] B. Nurina Sari, A. Primajaya, and J. H. Ronggowaluyo Teluk Jambe Karawang, “Penerapan Clustering Dbscan Untuk Pertanian Padi Di Kabupaten Karawang,” 2019. [Online]. Available: www.mapcoordinates.net/en.

[15] S. Zhu and P. Bao, “The use of Geographic Information System in the development and utilization of ancient local chronicles,” *Library Hi Tech*, vol. 33, no. 3, pp. 356–368, Sep. 2015.
https://doi: 10.1108/LHT-03-2015-0028.

[16] R. K. Dinata, S. Retno, and N. Hasdyna, “Minimization of the Number of Iterations in K-Medoids Clustering with Purity Algorithm,” *Revue d’Intelligence Artificielle*, vol. 35, no. 3, pp. 193–199, Jun. 2021.
https:// doi: 10.18280/RIA.350302.

[17] Luhung Mustika Budiharti and S. Sunendiari, “Pemodelan dan Pemetaan Jumlah Penderita Kusta di Jawa Barat dengan Regresi Binomial Negatif dan Flexibly Shaped Spatial Scan Statistic,” *Jurnal Riset Statistika*, vol. 1, no. 2, pp. 99–106, Dec. 2021.
https://doi: 10.29313/jrs.v1i2.409.

[18] T. N. Hikmaulida, M. N. Hayati, and S. Wahyuningsih, “Analisis Spasial Persebaran Dan Pemetaan Kerawanan Daerah Titik Panas Di Kalimantan Timur Menggunakan Spatial Pattern Analysis Dan Flexibly Shaped Spatial Scan Statistic,” 2020. [Online]
http://jurnal.fmipa.unmul.ac.id/index.php/ppj/indexHalaman| 20

[19] N. Ulinnhua and S. A. Sholihah, “Analisis Cluster Untuk Pemetaan Data Kasus Covid-19 Di Indonesia Menggunakan K -Means,” *Jurnal Matematika dan Statistika serta Aplikasinya (Jurnal MSA)*, vol. Vol. 09 No.2, pp. 14–18, 2021.

[20] B. Liu, Y. Xin, R. C. C. Cheung, and H. Yan, “GPU-based biclustering for microarray data analysis in neurocomputing,” *Neurocomputing*, vol. 134, pp. 239–246, Jun. 2014.
https:// doi: 10.1016/J.NEUCOM.2013.06.049.

[21] M. Douider, I. Amrani, T. Balenghien, A. Bennouna, and M. Abik, “Impact of recursive feature elimination with cross-validation in modeling the spatial distribution of three mosquito species in Morocco,” *Revue d’Intelligence Artificielle*, vol. 36, no. 6, pp. 855–862, Dec. 2022.
https://doi: 10.18280/RIA.360605.

[22] T. I. Hermanto and M. A. Sunandar, “Analisis Data Sebaran Penyakit Menggunakan Algoritma Density Based Spatial Clustering of Applications With Noise,” Nov. 2020. Accessed: Jan. 29, 2023. [Online]. Available: https://doi.org/10.33084/jsakti.v3i1.1775

[23] P. Verma, V. K. Awasthi, and S. K. Sahu, “A novel design of classification of coronary artery disease using deep learning and data mining algorithms,” *Revue d’Intelligence Artificielle*, vol. 35, no. 3, pp. 209–215, Jun. 2021.
https://doi: 10.18280/RIA.350304.

[24] M. Gaonkar and K. Sawant, "AutoEpsDBSCAN : DBSCAN with Eps Automatic for Large Dataset," 2013.

[25] N. Ahmed and A. Razak T, "An Overview of Various Improvements of DBSCAN Algorithm in Clustering Spatial Databases," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, no. 2, pp. 360–363, 2016.

[26] Y. Chen, "New Approaches for Calculating Moran's Index of Spatial Autocorrelation," *PLOS ONE*, vol. 8, no. 7, p. e68336, Jul. 2013. https://doi: 10.1371/journal.pone.0068336.