International Journal of GEOMATE, Aug., 2020, Vol.19, Issue 72, pp. 35 - 42
ISSN: 2186-2982 (P), 2186-2990 (O), Japan, DOI: https://doi.org/10.21660/2020.72.5599
Special Issue on Science, Engineering and Environment

GEOINFORMATICS OF TUBERCULOSIS (TB) DISEASE IN JAKARTA CITY INDONESIA

*Titin Siswantining[1], Niken Pratiwi Cahya Dewi Purwandani[2], Maria Hedwig Dewi Susilowati[2] and Adi Wibowo[2]

[1]Dept. of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Indonesia;
[2]Dept. of Geography, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Indonesia

*Corresponding Author, Received: 17 Aug. 2019, Revised: 20 Nov. 2019, Accepted: 24 Feb. 2020

ABSTRACT: Tuberculosis (TB) disease has still become the leading cause of morbidity and mortality in Indonesia. TB had a relationship between human and their environment, especially in an urban area. The urban area had the highest population and density, that is the way the accurate information about the urban environment of the TB area is critical. The research objective to representing characteristic spatial of TB patients lived in Jakarta year 2017. The research used a method using visual interpretation from superimposed spatial data. Spatial data collection from google earth data, household area (RW boundary), and TB patient addressed from the hospital. The total TB patients in Jakarta City year 2017 in total 114 from a hospital and distributed in 108 household areas (RW boundary). The result of geoinformatics of the urban environment, first building population, building density, vegetated covered, and distance from the main road. The TB area characteristic had two classified high and low, based on total patient TB disease in a household area (RW boundary). The research concluded of the geoinformatics of TB disease in Jakarta City year 2017 with highly classified is high building population and building density, lowest vegetated covered (high temperature), and near the main road. The research also uses the PAM algorithm to cluster the patient and resulted in four clusters with differ mostly in building density, building temperature, and the distance to the nearest river.

Keywords: Characteristic spatial, Geoinformatics, PAM clustering, TB Diseases

1. INTRODUCTION

Tuberculosis (TB) is one of the infectious diseases caused by Mycobacterium tuberculosis with the name of acid-resistant bacteria (BTA) [1]. In the year 2017, there have been 6.4 million official new TB cases reported to the WHO, this number increased since 2013 with a total of 5.8 million new cases that occurred in the year 2017, then the country with the top number of cases India (26%), Indonesia (11%), and Nigeria (9%) [2]. Indonesia recorded ten causes of death. Number one is an intensive stroke, number two is a heart attack, number three is diabetes mellitus, and the number forth is tuberculosis.

Based on the results of primary health research in 2013, DKI Jakarta is one of the provinces that has the highest TB prevalence rate in Indonesia because it occupies the second position of the equivalent to Papua is 0.6 [3]. The case of a positive BTA pulmonary tuberculosis or bacteria that is red and rod-shaped in Indonesia in 2016, according to the most age group found in the age group 45-54 years, and the record that the sufferer occurs at a productive age Compared to non-productive age. In the year 2016, DKI Jakarta recorded the most significant Case notification rate number of 241 cases, and the success rate of treatment is only 73% [4].

The tools for dealing with geospatial information have primarily developed with the context of an emerging discipline, named Geoinformatics, Geomatics, or Geographic (sometimes Geospatial) Information Sciences [5]. The term of Geomatics is derived from the French word Ge´omatique, coined by the French photogrammetric Dubuisson and widely used in North America [6]. Geoinformatics seems to be popular in Europe, whereas Geographic Information Science usually indicates a Geography background [7], [1]. In the following, the term Geoinformatics will use because it emphasizes a formal scientific approach to handle geoinformation, which Geomatics does not imply. Tuberculosis (TB) disease has still become the leading cause of morbidity and mortality in Indonesia.

A spatial approach can examine the geosphere phenomenon of the complexity of the symptoms with the concept of a regional paradigm [8]. The territory is a characteristic similarity in a part of the Earth's surface [9]. The spatial approach in each region can cope with the health problems that considered as the top priority, with the resources being able to be used more effectively [10]. TB had a relationship between human and their environment, especially in the urban area with the highest population and density.

TB had a relationship between human and their environment, especially in an urban area. The urban area had the highest population and density, that is the way the accurate information about the urban environment of the TB area is critical. The research objective to representing characteristic spatial of TB patients lived in Jakarta year 2017.

2. METHODS

The research conduct in Jakarta Province in July until November 2018. The unit of analysis used is the point of the sufferer of TB and the household area administrative boundary (RW boundary), whose recorded as tuberculosis disease in the year 2017. The local characteristic variables in the TB-sufferer or RW in the study include building density, air temperature, and distance from the river, and the quality of the settlement. Data on residential environmental conditions from Google Earth and surveys. The data used includes TB disease sufferers of 114 patients with tuberculosis disease from the National Center General Hospital of Dr. Cipto Mangunkusumo in 2017 [11, 12, 13, 14, 15, 16].

The research used visual interpretation from spatial data collection from google earth data, RW boundary, and TB patient addressed from the hospital. The total TB patients in Jakarta City year 2017 in total 114 from a hospital and distributed in 108 household areas (RW boundary).

This research also uses a clustering algorithm called K-medoids. K-medoids or better known as partitioning around medoids (PAM) is a clustering algorithm that works almost like K-mean clustering. K-medoids work by partitioning data to minimize the distance between point labeled to be in a cluster and a point (should be in data) designated as the center (medoids) of the cluster. The algorithm works as follow:

1. Choose k of n data points as the medoids that minimize the cost function
2. Associate each data point to the closest medoid
3. For each medoid m and each non-medoid data point l, search for the best l such that if we swap m and l, the cost function that most. If none, the algorithm finished. Do this until no changes in the cost function.

With the flow chart shown in Fig 1. One advantage of using PAM as the clustering method is that it is intuitive, more robust to noise dan outliers compared to k-means, and it produces a typical data point for each cluster while the disadvantages of using PAM is time-consuming and computer-intensive [17].

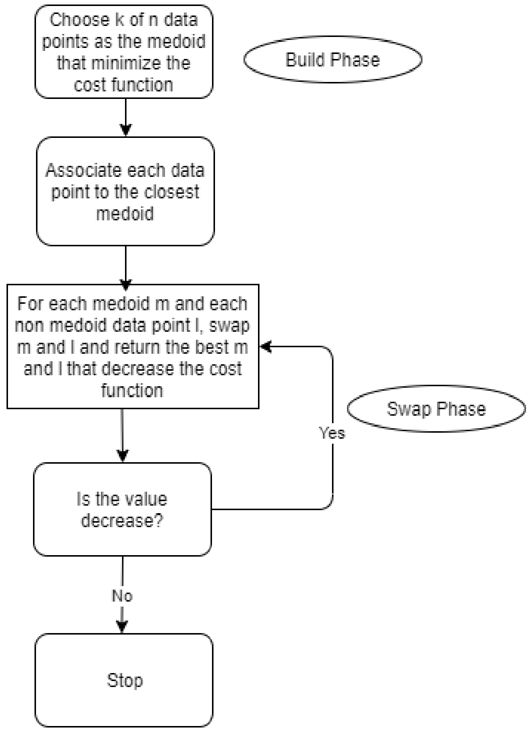

Fig. 1 Flow chart of the PAM algorithm

A cost function that used to cluster mixed data (categorical and numerical data) is Gower distance. Suppose that there are p features. Gower distance for different two data points is calculated as the average of partial dissimilarities ($\boldsymbol{d_{ij}^{(f)}}$) that is

$$\boldsymbol{d(i,j) = \frac{1}{p}\sum_{i=1}^{p} d_{ij}^{(f)}}.$$

Partial dissimilarity is calculated based on the type of feature time-consuming.

- For a numerical feature *f*, partial dissimilarity is the ratio between the absolute difference between the data points and the maximum range observed across all data points that is

$$d_{ij}^{(f)} = \frac{|x_{if} - x_{jf}|}{\max_{i,j}|x_{if} - x_{jf}|}$$

- For a qualitative feature *f*, partial dissimilarity equals 1 only if the data points have the same value of *f*. Zero otherwise.

Step 1 in the PAM algorithm work as follows (this step is called BUILD while step 3 is called SWAP):

Denote S as the set of the selected point, U is the set of unselected point.

For each data point p, we maintain two numbers:

- D_p the dissimilarity between p and the closest object in S
- E_p the dissimilarity between p and the second closest object in S.

Notice that $\boldsymbol{D_p} \leq \boldsymbol{E_p}$ and we have $\boldsymbol{p} \in \boldsymbol{S}$ if and only if $\boldsymbol{D_p} = \boldsymbol{0}$. The BUILD phase entails the following steps:

1. Initialize S by adding to it a data point for which the sum of the distance to all other data points is minimal
2. Consider a data point $i \in U$as a candidate for inclusion into S
3. For a data point, $j \in U - \{i\}$ compute D_j.
4. If $D_j > d(i,j)$, data point j will contribute to the decision to select object i. Let $C_{ji} = \max\{D_j - d(j,i), 0\}$
5. Compute the total gain obtained by adding i to S as $g_I = \sum_{j \in U} C_{ji}$.
6. Choose the object i that maximizes g_i, we reassign S as $S \cup \{i\}$ and $U = U - \{i\}$.
7. Do 1-6 until $|S| = k$.

Here, the optimal number of clusters is determined according to the elbow method in the SWAP phase.

3. RESULT AND DISCUSSIONS

3.1. Patient with TB Disease

Patients with TB disease in Jakarta in 2017 amounted to 114 sufferers, which scattered in 35 sub-districts, 80 villages, and 108 pillars.

Table 1 Number of TB Patient

No	Administrative	Patient With TB	(%)
1	North Jakarta	12	10.5
2	Central Jakarta	34	29.8
3	South Jakarta	20	17.6
4	West Jakarta	12	10.5
5	East Jakarta	36	31.6
	Total	114	100.0

Sources: Data processing

The location of the people of TB is seen by spatial many occur in central Jakarta, South Jakarta, and East Jakarta. (see fig. 2 and Table 1). Based on the data of TB patients, there are three characteristics of sufferers, including gender, age, and nutritional status.

Table 2 Number of TB Suffers based on Gender

Gender	Number	Percentage (%)
Male	79	69
Female	35	31
Total	114	100

Source: data processing

Patients with male tuberculosis disease amounted to 79 sufferers or 69%, and female-sex sufferers amounted to 35 or 31% (Table 2). Based on age, patients with tuberculosis disease who are productive (15-64 years) amount to 87 sufferers or 77%, and non-productive age (< 15 and > 64 years) amounted to 27 sufferers or 23%. Based on nutritional status, patients with tuberculosis disease who have a lean nutritional status (BMI value < 17-18.4) amounted to 36 sufferers or 32%, normal nutritional status (BMI value 18.5-25) amounted to 63 or 55%, and fat nutritional status (BMI value 25-27) amounted to 15 or 13%.

3.2 Spatial Distribution of TB Sufferer

Areas with TB disease in DKI Jakarta year 2017 amounted to 108, which includes 114 people with TB (see fig. 3). The lowest area of TB sufferers or 1 TB sufferer per RW as much as 102 regions are in an RW that spreads from the edge to the city center. While in the height of TB sufferers or 2 TB sufferers per RW as many as six regions located in the central to the east side of the city, which precisely located in the village of Malacca Sari, Paseban, Rawasari, Senen, and Tanah Tinggi. Clustering based on the spatial distribution of TB patients and resulted as follows in Figure 4.

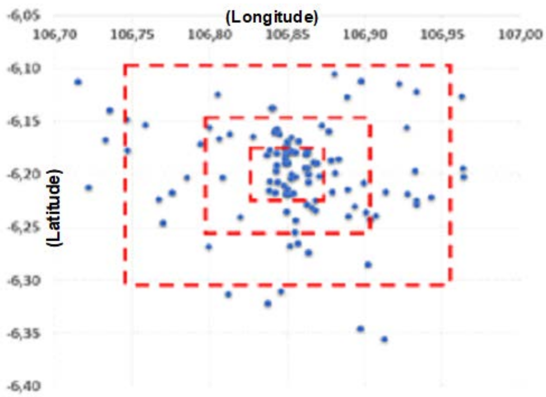

Fig.4 Cluster of TB Disease in Jakarta

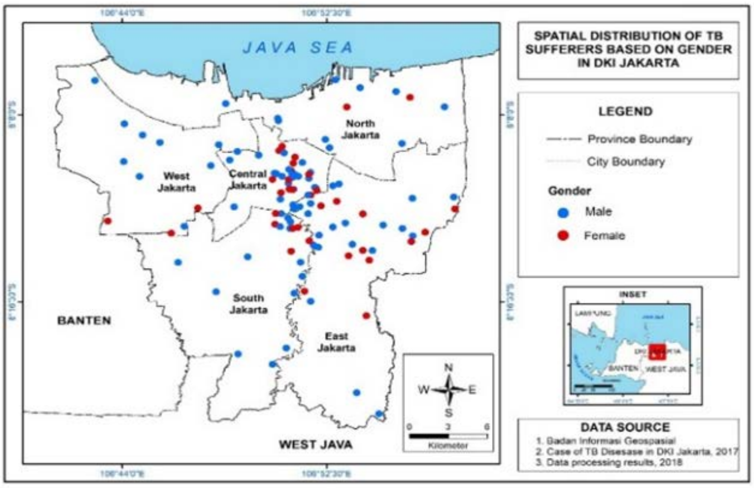

Fig.2 Spatial distribution of TB Suffers in DKI Jakarta year 2017

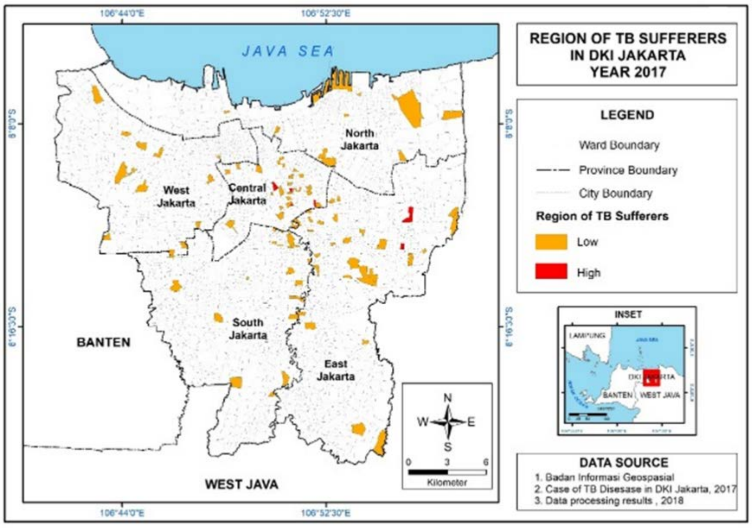

Fig.3 Spatial Distribution of TB Sufferers regions in DKI Jakarta year 2017

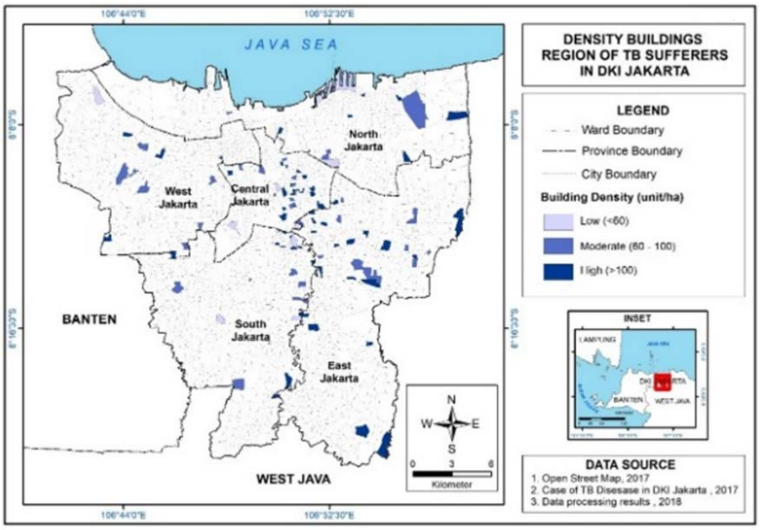

Fig. 5 Building Density of TB Suffer in DKI Jakarta

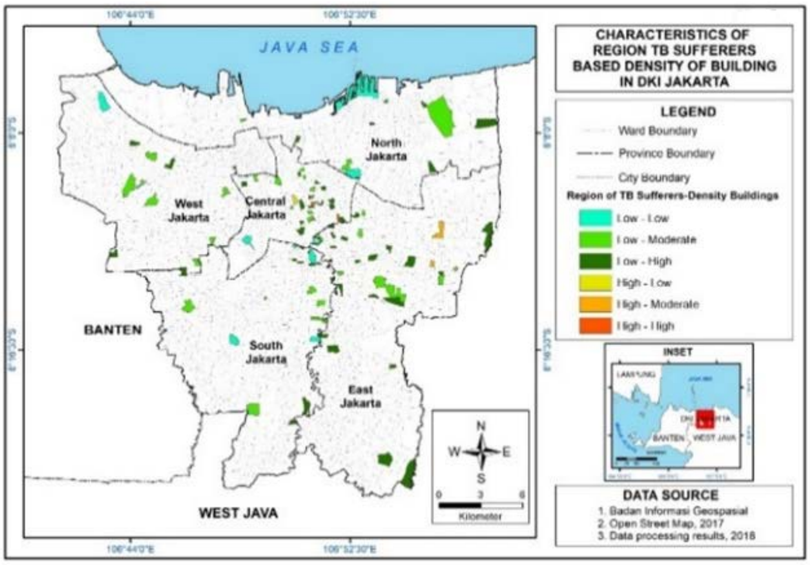

Fig. 6 The characteristic of TB sufferer region based on building density in DKI Jakarta

3.3 Spatial Distribution of Building Density

Building Density obtained from the Intersect results of Open Street Map building with the region of TB sufferers, then classified into three, namely low density (< 60 units/ha), Moderate (60-100 units/ha), and high (> 100 units/ha). Based on the building density map (see fig. 5), the characteristics of areas with a high density of buildings located in the central and eastern areas. In the downtown area, which TB sufferers have a group of the pattern, as for the eastern region has a pattern spread from east to southeast.

The density of building Data with TB region done cross-tabulation, which produces six classifications, so it is symbolized with six different colors to see the difference of area characteristic of sufferer based on building density (see fig. 6). Based on the explanation, it shown that TB sufferers not only occur at the high density of buildings can even occur at the density of low and medium buildings.

3.4 Environment Quality

The first quality of settlement in parameters is the layout of the building in an RW. The Data shows that people with TB has the orientation of different buildings that are good, moderate and bad so that the condition of the building layout can be infected by TB disease (see fig. 7).

The people with TB disease have the condition of the building with cold, no air circulation, no sun entry to the building so that conditions will be infected by TB disease (see fig. 8)

Second is the protective tree; it has seen that the low or high of TB sufferers have a protective tree condition or the existence of harmful trees or < 25% (see fig. 9). There was tree cover in the RW, which is assumed to increase the temperature of the air and cause the air cycle is not optimal in the absorption of pollutants. When the air temperature rises, the condition inversely proportional to the humidity of the air decreases, making it easier for germs to breed.

Fig. 8 Building layout of TB patient

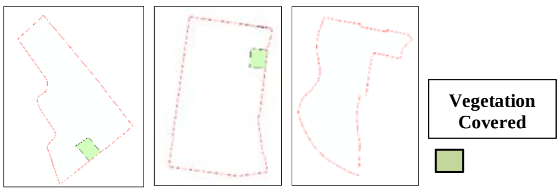

Fig. 9 The tree-covered (green color) in the household area (RW boundaries in red color)

The third is the width of the road; it sees that TB can occur in a residential area that has a category of good or easy to go through two cars, medium or only traversed one car, or harmful or without a car can pass. However, when viewed based on a high area of TB sufferers, it tends to occur in areas that have a narrow or steep road width traversed with four-wheeled vehicles, and they have a moderate to high density (see fig. 10).

Fig. 7 Building (grey color) density based on the household area

Fig. 10 The road width in the household area (RW boundaries)

Fourth, based on the location of the settlement against the source of pollution and activities, poor settlement environment conditions following the high area of TB sufferers, it is explained that the closer to the source of pollution, the main road will facilitate the occurrence of transmission in TB disease. The result appropriate with previous research by Fitriana [18], which explains that the activity center or source of pollution is a place to interact to facilitate the transmission of TB (see fig. 11).

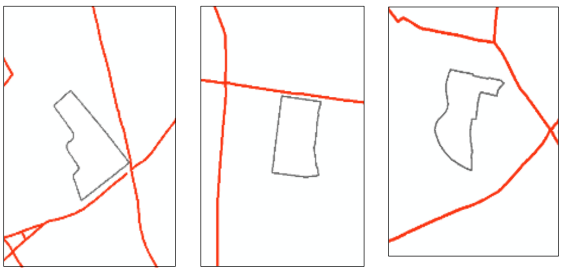

Fig. 11 The highway road (red color) near the household area (RW boundaries in grey color)

Lastly, the health service parameters appear that from the six regions can still be affordable to health facilities because it is in the classification well to moderate. DKI Jakarta is certainly a metropolitan city that has health facilities almost spread evenly so that in TB, people who are close to health services can also occur TB sufferers. Because of not only the treatment factors available but also the needed routine in medicine for the sufferer.

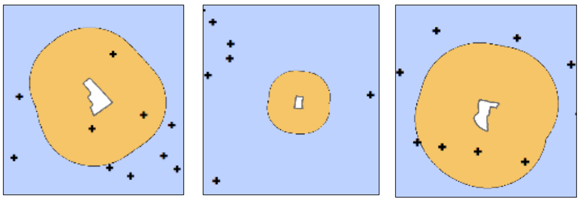

Fig. 12 The health facilities (black point) around (brown color) the household area (RW boundaries area with white color)

Based on the explanation, spatial uniqueness that occurred in 6 TB sufferers turned out to have a common characteristic of the region that is irregularities in the layout of buildings, and road width tends to narrow, the existence of protective trees and close to the main road as a place of activity or activities (see fig 12).

Using the PAM algorithm with Gower distance, we make two cluster scheme which based on the variable type of age and BMI value. The first cluster scheme is using numerical value age and BMI. From here, it was found that the patient tends to make four clusters which are

- Cluster 1 (29 patients) is dominant by males with ages ranging from 45 to 65 years old. These patients have BMI around 20.3 and live in high building population and density area with medium temperature and medium distance to river.

- Cluster 2 (36 patients) is dominant by males with ages ranging from 45 to 64 years old. These patients have BMI around 21.51 and live in medium building population and density area with median temperature and relatively near to the river compared to another cluster.
- Cluster 3 (25 patients) is dominant by males with ages ranging from 41 to 56 years old. These patients have BMI of around 20.37 and live in high building population and density area with the highest temperature and farthest distance to river.
- Cluster 4 (24 patients) is dominant by a female with ages ranging from 29 to 52.5 years old. These patients have BMI around 20.18 and live in high building population and density area with medium temperature and low distance to river.

The second scheme uses the categorical value of age and BMI with the resulted cluster described as follow:

- Cluster 1 (32 patients) is dominant by males with a productive age (75%). These patients have a medium value of BMI (18.5 to 25.0). These patients live in high building populations and density areas with medium temperature and medium distance to river.
- Cluster 2 (35 patients) is dominant by males with a productive age (77%). These patients have a medium value of BMI (18.5 to 25.0) and live in medium building population and density area with median temperature and relatively near to river compared to another cluster.
- Cluster 3 (24 patients) is dominant by males with a productive age (75%). These patients have a medium value of BMI (18.5 to 25.0) and live in high building population and density area with the highest temperature and farthest distance to river.
- Cluster 4 (23 patients) is dominant by a female with a productive age (78%). These patients have a low value of BMI (below 18.4) and live in high building population and density area with medium temperature and low distance to a river
-

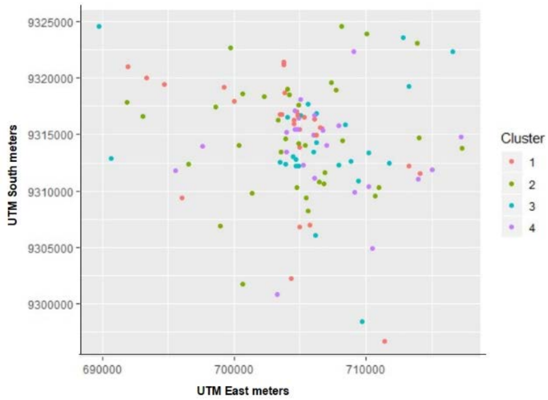

Fig. 13 The spatial distribution of the first cluster scheme

From the above result, it can seem that the cluster method, either using the first scheme or the second scheme, resulted in a consistent cluster of patients. Furthermore, to see the spatial distribution of the TB patient, the result plot the distribution for the first scheme and get the following result.

From Fig. 13, the researchers concluded that the spatial distribution for each cluster spread across the area. Also, from the second cluster scheme (Fig. 14), the research gets the following spatial distribution of TB patients.

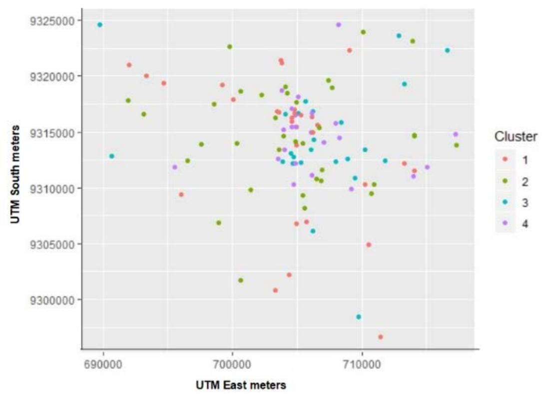

Fig. 14 The spatial distribution of the second cluster scheme

Based on Fig. 13 and Fig. 14, the researchers concluded that the consistency of the clustering also happens in the spatial distribution of the TB patients. Thus, it can explain the PAM algorithm is consistent, whether if use the numerical value of age and BMI or the categorical value of age and BMI.

4. CONCLUSION

The research concluded that the geoinformatics of TB disease in Jakarta City year 2017 with highly classified is high building population and building density, lowest vegetated covered, and near the main road. The research also concludes that the patient tends to make four clusters that affected mostly by the building temperature, building density, and the distance to the river. For future work, the clustering algorithm used can be changed to ROCK clustering, which is well known for the robustness, but it takes a longer time than the PAM algorithm.

5. ACKNOWLEDGMENTS

The research support by the Universitas Indonesia Research Grant with the contract number was NKB-0667/UN2.R3.1/HKP.05.00/2019.

6. REFERENCES

[1] Data Center and Information Ministry of Health Republik Indonesia, Tuberculosis. Ministry of Health Republic Indonesia, 2018.
[2] World Health Organization, Global Tuberculosis Report, 2018.
[3] The Research Agency and Health Development, Indonesian Ministry of Health, Report Research of Basic Health (RISKESDAS), 2013.
[4] Ministry of Health Republik Indonesia, Data and Information Indonesia Health Profile 2016, Indonesia Ministry of Health Data and Information Center, 2017.
[5] Chandarasekaran S.K., and Arivarignan G., Disease mapping using mixture distribution, Indian Journal of Medical Research, Vol. 123, Issue 6, 2006, pp. 788-798.
[6] Ehler M., Geoinformatics and digital earth initiatives: a German perspective, International Journal of Digital Earth, Vol. 1, Issue 1, 2008, pp. 17-30.
[7] Goodchild M.F., Geographical Information Science, International Journal of Geographical Information Systems, Volume 6, Issue 1, 1992, pp. 31-45.
[8] Yunus H.S., Cotemporary Regional Research Methodology, Pustaka pelajar Yogjakarta, 2010.
[9] Sumaatmadja N., Geography Study: An Spatial Approach and Analysis, Bandung ALumni, 1981.
[10] Fauzan A., Achmadi U.F. and Susanna D., Spatial Analysis of Positive BTA Lung Tuberculosis in Sukabumi City, Universitas Indonesia, 2014.
[11] Siswantining T., Maylinda A., Kamelia T. and Bustamam A., Risk Factors for Resistance Tuberculosis, AIP Conference Proceedings 2023, 020224, 2018.
[12] Anisa I.R., Siswantining T., Kamelia T. and Sarwinda D., Risk Prediction Model Occurrence of Obstructive Sleep Apnea (OSA) Risk in Chronic Obstructive Pulmonary Disease patient, Case Study: Cipto Mangunkusumo Hospital, IOP Conf. Series: Journal of Physics: Conf. Series 1108,012076, 2018.
[13] Aristiawati K., Siswantining T., Sarwinda D., and Soemartojo S.M., Missing values imputation based on fuzzy C-Means algorithm for classification of chronic obstructive pulmonary disease (COPD). AIP Conference Proceedings 2192, 060003, 2019.
[14] Fajar I., Siswantining T., Soemartojo S.M., The estimated proportion of chronic disease sufferer in Duren Sawit district, East Jakarta, using hierarchical bayes method in small area estimation (SAE), Journal of Physics: Conference Series, Vol. 1321, No. 2, 2019, 022061,.
[15] Vitasari D.N., Siswantining T., Kamelia T., Identification of factor affecting atrial fibrillation in a patient with risk of obstructive sleep apnea at Rumah Sakit dr. Cipto Mangunkusumo using decision tree method, Journal of Physics: Conference Series, Vol. 1321, No. 2, 2019, 022108.
[16] Puspitaningtyas F.E., Siswantining T., Kamelia T., Analysis of obstructive sleep apnea, diabetes mellitus type 2, and prediabetes at dr. Cipto Mangunkusumo hospital using partial least squares, Journal of Physics: Conference Series, Vol. 1321, No. 2, 2019, 022062,.
[17] Rosalia D.C., Bustamam A. and Siswantining T., Implementation of Spectral Clustering with Partitioning Around Medoids (PAM) Algorithm on Microarray Data of Carcinoma, AIP Conference Proceedings 1825, 2017, 020007,.
[18] Fitriana N., Saraswati E. and Widayani P., Remote sensing and Geographic Informations System Applied for mapping Tuberculosis (TB) vulnerability in Imogiri District, Bantul Regency, Yogyakarta, Bumi Indonesia Journal, Vol. 3, Issue 2, 2014, pp. 110-120.