Clustering tuberculosis in children

by Purwanto Purwanto

Submission date: 29-Mar-2020 12:07AM (UTC+0700)
Submission ID: 1283975538
File name: Int_Proc_8_Clustering_tuberculosis_in_children_using_KMeans.pdf (811.42K)
Word count: 3017
Character count: 15719

Clustering tuberculosis in children using K-Means based on geographic information system

Cite as: AIP Conference Proceedings **2114**, 060012 (2019); https://doi.org/10.1063/1.5112483
Published Online: 26 June 2019

Ratih Sari Wardani, Purwanto, Sayono, and Aditya Paramananda

Clustering Tuberculosis in Children Using K-Means based on Geographic Information System

Ratih Sari Wardani[1,a)], Purwanto[2, b)], Sayono[3, c)] and Aditya Paramananda[4, d)]

[1, 3]*Universitas Muhammadiyah Semarang, Semarang, Central Java, Indonesia*
[2, 4]*Universitas Dian Nuswantoro, Semarang, Central Java, Indonesia*

[a)] Corresponding author: ratihsw@unimus.ac.id
[b)] purwanto@dsn.dinus.ac.id
[c)] say.epid@ unimus.ac.id
[d)] paramananda.aditya@gmail.com,

Abstract. The prevalence of tuberculosis (TBC) in children tends to increase and become a serious public health problem. The higher case of TBC in children requires comprehensive surveillance. Surveillance can be carried out by clustering the cases of TBC in children based on region. One of the clustering methods that can be used is the K-Means method. Furthermore, Geographic Information Systems (GIS) has been successfully used in health fields and have the advantages in the processing and presentation of spatial and non-spatial information. The present research aims to classify the cases of TBC in children using K-Means clustering and identify the distribution patterns using GIS. Based on the analysis, the spatial pattern of the distribution of TBC in children can be categorized into: areas with high prevalence, areas with moderate prevalence, and areas with low prevalence. The result can be used to help decision making in controlling the cases of TBC in children.

INTRODUCTION

TBC is an infectious disease that attacks the lungs and other organs caused by the Mycobacterium Tuberculosis. It has been a serious global health problem, including in Indonesia. The World Health Organization (WHO) reports the high prevalence of TBC and its tendency to increase as indicated by 8.3 million new cases in 2000 to 9.7 million in 2011 and 10.4 million in 2016. From this data, about 1 million patients diagnosed with TBS are children [1].

Indonesia is the world's second country with highest number of patients suffering from TBC in which in 2015, the case of TBC in children was approximately 75 per 100,000 populations. Subsequently, it decreased to 60 per 100,000 populations in 2016. In 2012-2016, the case was 8.21%-9.04% for children aged 0-4 and 15.80%-16.19% for children aged 5-14, of which about 3% were also diagnosed with HIV/AIDS Tuberculosis [2].

Likewise, the prevalence of TBC in children in the Province of Central Java, Indonesia, also shows an increasing trend. In 2014, there were 1977 cases or 6.63% of the national total cases. It increased up to 2975 or 7.5% in 2015, then decreased to 2585 or 6.47% in 2016 [3]. Similar situation was also depicted in Semarang, in which the number of the case reached 433 in 2014, 405 in 2015, and 492 in 2016 [4].

The increasing number of TBC case in children implies the urgency for more comprehensive treatment efforts, since health problems will experience changes based on time, as well as differences in places. It will cause variability of problem between regions (spatial) that will have specificity in accordance with the region [5]. The expansion of this case is categorized by the region with clustering method. It is a part of data mining which main goal is to set data or objects into groups (clusters) so that each group contains almost similar data. The clustering method commonly used to map or classify an object is the K-Means algorithm [6,7]. It has been widely used to set cluster of disease, such as in India (for diabetics) [8] and acute respiratory infections [9], and there is a potential to use it for clustering the case of TBC [10]. Health mapping can be performed using GIS by entailing spatial and non-

Exploring Resources, Process and Design for Sustainable Urban Development
AIP Conf. Proc. 2114, 060012-1–060012-9; https://doi.org/10.1063/1.5112483
Published by AIP Publishing. 978-0-7354-1850-9/$30.00

spatial data processing capabilities, and presentation of health information. It is an important tool to determine spatial-based health risk factors, public health trends and vulnerabilities [11]. In general, GIS has several advantages in relation with its features in data collection, data management, data analysis as well as data report and visualization [12]. Combining clustering methods in data mining with GIS is also a challenge for researchers. The combination of methods is expected to help the process of grouping objects in accordance with the existing spatial patterns in order to help the control of tuberculosis through TBC surveillance activities.

A research carried out in Bandar Lampung has employed TBC-GIS and grouped two clusters based on space-time statistics. They are the cluster of areas with high population density and the cluster of a high proportion of poor families and housing [13]. Likewise, a study of TBC in Thailand was set in two clusters, namely Mueang Si Sa Ket and Kukhan [14]. It reveals that spatial patterns were identified as risk factors related to TBC cases, including spatial, socioeconomic and environmental factors [15].

The present study is necessary to classify the prevalence of TBC in children using K-Means clustering and identify the case distribution patterns using GIS, so that it can be used to help decision-making in controlling the cases of TBC in children.

METHODOLOGY

Data Collection

The data collection in this study was done to obtain the secondary data related to TBC from the TBC register of the Health Office of Semarang city on 2014-2017. As many as 1,965 cases were examined.

Production of Digital Mapping

The production of a digital map of Semarang with the administrative boundary of the sub-district with a scale of 1: 25000 was carried out. The map is equipped with attributes in the form of data on the cases of TBC in children in Semarang (2014-2017).

Categorization by Using Clustering K-Means Method

The K-Means clustering method aims to group the data set to a cluster K. The steps of K-Means clustering algorithm are as follows [16]:

1. Select number of clusters (K). This study uses 3 (three) groups of cluster, namely: low, moderate and high.
2. Determine cluster centers (centroid).
3. Calculate the distance measure (distance similarity) of each distance to every centroid. For the process, this study used Euclidean distance formula.
4. Assign the data or objects based on the nearest distance. If there is data that moving from the groups then go to step 2, otherwise the process is considered to be complete.

The algorithm of K-Means clustering is shown in the flowchart below.

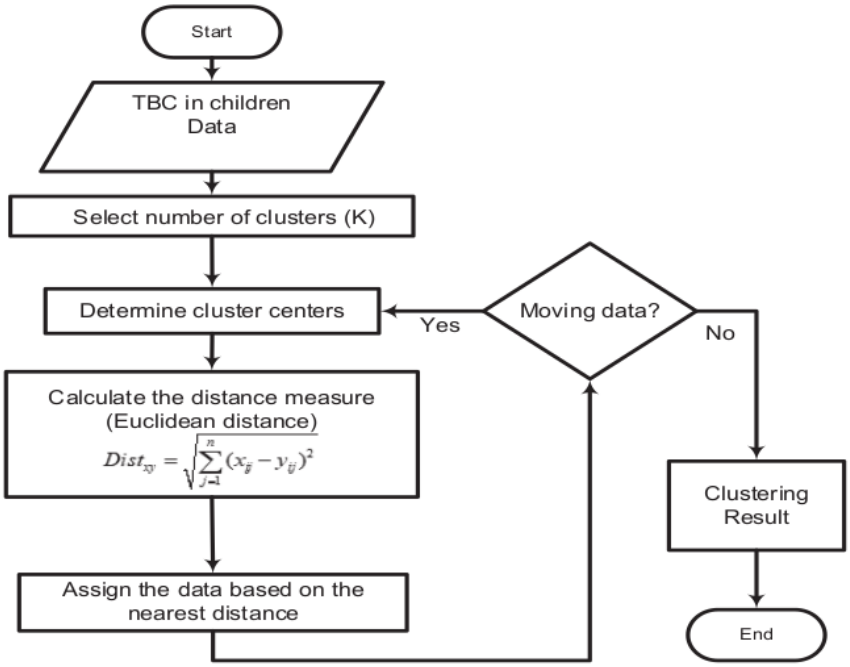

FIGURE 1. Algorithm of K-Means clustering

Combination of Clustering K-Means Method with GIS

In the next step, a digital map with the sub-district administrative boundaries was carried out. The results of clustering of TBC in children using K-Means method will be used to obtain a spatial map of the case distribution in Semarang.

RESULTS AND DISCUSSION

The Case of TBC in Children in Semarang

The data of TBC case in children in Semarang was obtained from various health services, such as public health centers, hospitals and clinics, in 2014 - 2017 with total 1,965 cases. The finding of new cases of TBC in children was set into pulmonary tuberculosis with bacteriologically confirmed, pulmonary TBC with clinically diagnosed and pulmonary extra tuberculosis. The majority of new cases of tuberculosis in children is a clinical diagnose as 89% and bacteriological confirmed as 4% as shown in Fig. 2.

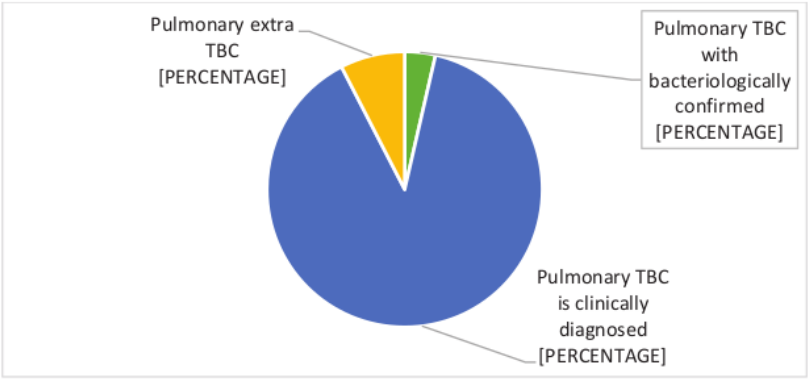

FIGURE 2. The diagnose result of new case of TBC in children in Semarang (2014 – 2017)

The Case of TBC in Children in Semarang (2014 – 2017) Based on Sex

The case of TBC in Children based on sex, Fig. 3 shows that the majority of children in Semarang suffering TBC is male (52%).

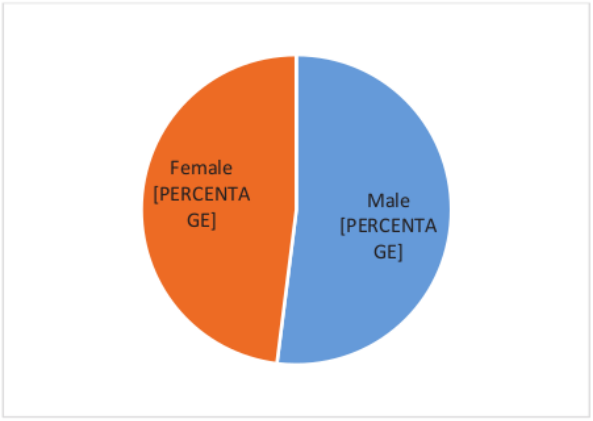

FIGURE 3. The frequency distribution of new case of TBC in children in Semarang (2014-2017) based on sex

The Case of TBC in Children in Semarang (2014 – 2017) based on Age

The new case of TBC in children based on age is shown in Fig. 4. The majority of children suffering from TBC are in the group age of 0-4 years old (60%).

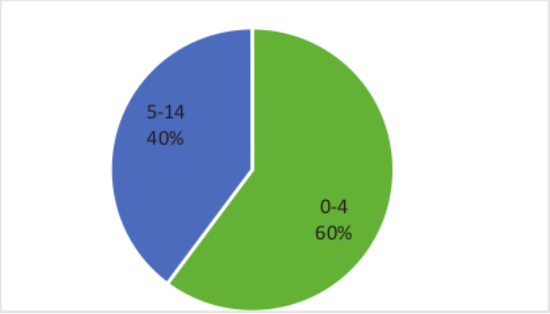

FIGURE 4. The frequency distribution of new case of TBC in children in Semarang (2014-2017) based on age

The Case of TBC in Children in Semarang (2014 – 2017) based on Area

The cases of TBC in children in Semarang (2014 – 2017) were grouped into 16 sub-districts according to the district administrative boundaries. Semarang Tengah has the highest number with 458 cases and Tugu has the lowest number with 17 cases. The detail is presented in Fig. 5.

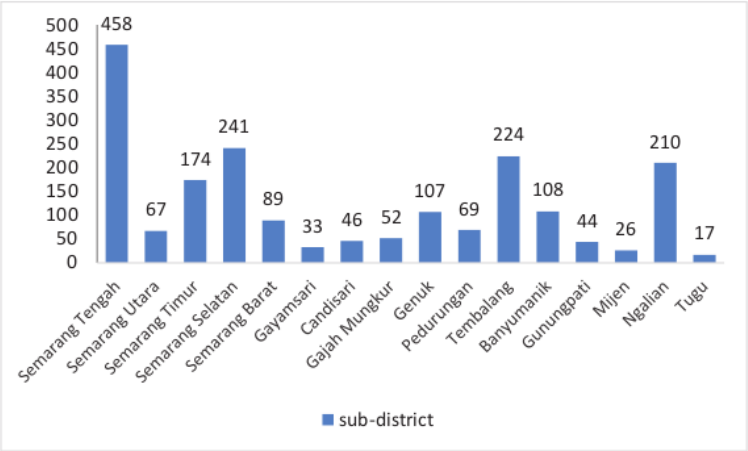

FIGURE 5. The frequency distribution of TBC in children in Semarang (2014 – 2017) based on the sub-district

The Set of TBC in Children using K-Means Clustering Method

The grouping of TBC in children using K-Means clustering method was carried out based on steps in the K-Means algorithm. The final result of the grouping is shown in Table 1.

TABLE 1. Grouping of TBC in Children by K-Means Clustering Method

No	Sub-District	0-4	5-14	C1	C2	C3	Nearest Distance
1	Semarang Tengah	281	177	0.00	228.30	298.71	0.000
2	Semarang Utara	34	33	285.91	57.65	16.73	16.730
3	Semarang Timur	115	59	203.67	30.37	95.95	30.370
4	Semarang Selatan	116	125	173.00	71.71	138.78	71.708
5	Semarang Barat	54	35	267.76	39.88	30.99	30.994
6	Gayamsari	21	12	307.94	80.04	9.23	9.232
7	Candisari	26	20	299.46	71.29	3.91	3.908
8	Gajah Mungkur	30	22	295.00	66.86	5.04	5.037
9	Genuk	63	44	255.37	27.23	43.59	27.234
10	Pedurungan	39	30	283.15	54.92	16.49	16.495
11	Tembalang	158	66	165.68	73.54	138.25	73.539
12	Banyumanik	46	62	261.63	38.70	48.09	38.696
13	Gunungpati	25	19	300.83	72.68	4.11	4.110
14	Mijen	19	7	312.32	84.70	14.01	14.011
15	Ngalian	146	64	176.05	61.43	126.34	61.433
16	Tugu	10	7	319.91	91.86	21.24	21.244

Since there has been no change in the pattern, the model for the prevalence of TBC case in Semarang (2014- 2017) gives result of 3 (three) groups. The groups are inputted to the GIS and the result is shown in Fig. 5.

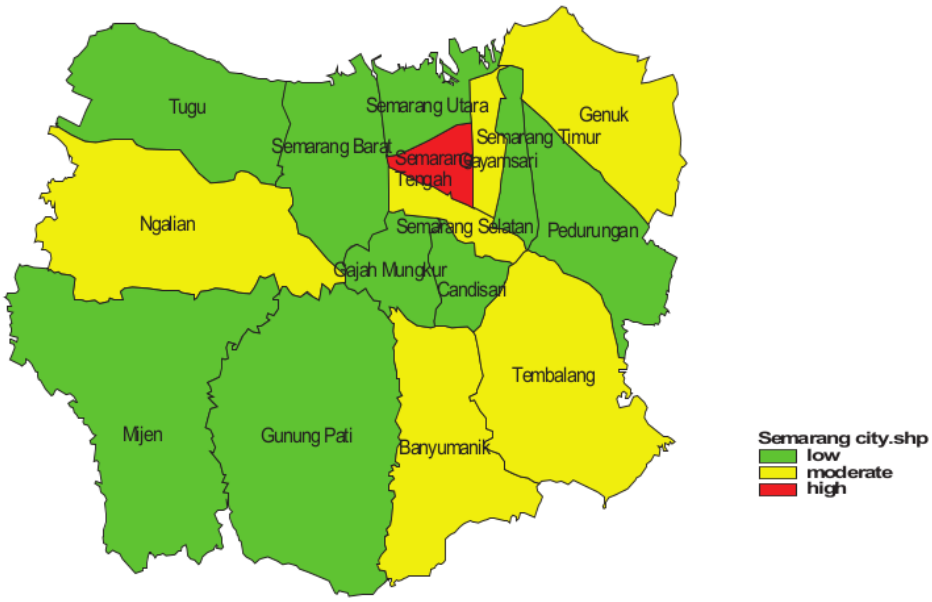

FIGURE 6. The Spatial Map of TBC in Children in Semarang (2014 – 2017) with K-Means clustering.

The figure above shows the result of the spatial pattern model using K-Means clustering that can be elucidated as follows:

1. Area with high prevalence (red) is Semarang Tengah.
2. Areas with moderate prevalence (yellow) are Ngalian, Semarang Selatan, Semarang Timur, Genuk, Banyumanik, and Tembalang.
3. Areas with low prevalence (green) are Tugu, Semarang Barat, Semarang Utara, Gayamsari, Pedurungan, Gajah Mungkur, Candisari, Mijen, and Gunung Pati.

The data used to conduct TBC in children clustering is based on the case findings in the region, the sources of the data are various health services, including Puskesmas, clinics and hospitals. The findings of the present study also indicate that moderate and high prevalence are dominant in areas where there are hospitals, both government hospitals and private hospitals.

In addition, the highest prevalence of the TBC cases is identified in Semarang Tengah, which is a dense residential area. Semarang Selatan and Gayamsari are those with moderate prevalence of TBC case. These areas are densely populated and used for settlement, while Ngalian, Tembalang, Banyumanik and Genuk are alternative areas for new settlement in Semarang as shown in Fig. 6. The finding of the present study is similar to research carried out in Lampung, which found out high distribution or cluster in areas with high population density [13].

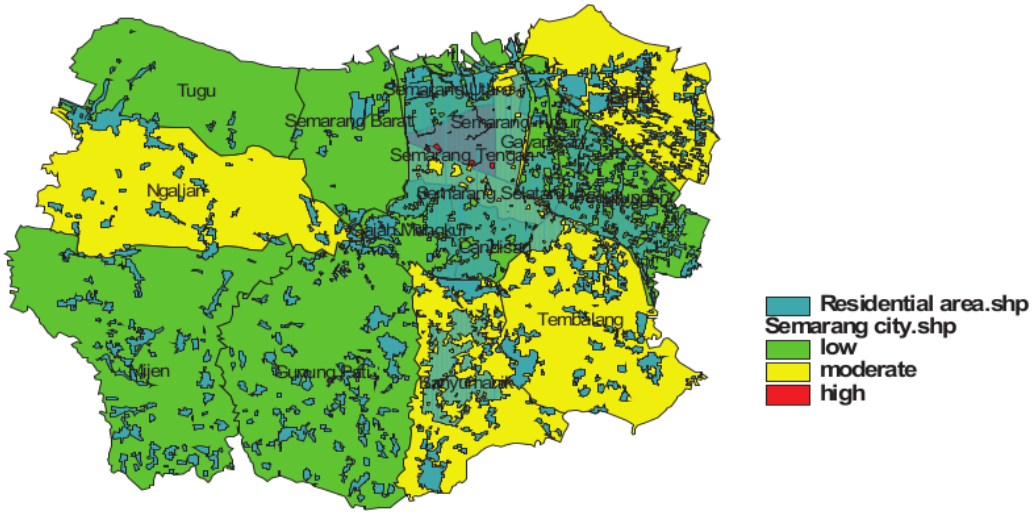

FIGURE 7. Map of TBC case in children overlay with map of residence.

The high prevalence of TBC in children showed high transmission of adult tuberculosis so it seems that the trend will remain if the TBC cases in adult are left untreated [17]. In addition, research in South Africa stated that childhood infection and disease are quantitatively linked to infectious TB prevalence among adults in an immediate social network. Childhood infection should be monitored in high-burden settings as a marker of the on-going TB transmission [18]. The results of spatial based TB clustering show that combining spatial data visualization using GIS and the use of statistical methods (K-Means clustering method) can be used to determine the observed patterns of disease observed spreading randomly or forming a cluster [19]. Identification of clustering helps in knowing early on the outbreak of infectious diseases, spatial analysis is also useful for detecting areas with a high risk of TBC, so that they can carry out best practices for TBC prevention and control [20].

CONCLUSION

The present study generates a spatial pattern model using K-Means clustering in Semarang for mapping TBC cases in children in 2014-2017. The result of clustering consists of three areas, which are specified as follows:

1. Area with high prevalence of TBC case as demonstrated by red zone is Semarang Tengah,
2. Areas with moderate prevalence as demonstrated by yellow zone are Ngalian, Semarang Selatan, Semarang Timur, Genuk, Banyumanik and Tembalang,
3. Areas with low prevalence as demonstrated by green zone are Tugu, Semarang Barat, Semarang Utara, Gayamsari, Pedurungan, Gajah Mungkur, Candisari, Mijen and Gunung Pati.

ACKNOWLEDGMENT

This paper is the result of a study under research grant supported by the Ministry of Research, Technology and Higher Education of the Republic of Indonesia in 2018.

REFERENCES

[1] WHO, *Global Tuberculosis Report 2018*. World Health Organization, Geneva, 2018). pp. 13-199 available at: https: www.who.int.
[2] Ministry of health, Indonesian Health Profile 2017, Jakarta, 2017, p.25. available at : http://www.depkes.go.id
[3] Central Java Provincial Health Office, Health Profile of Central Java Province in 2015, Semarang, 2016. p. 18 https://www.dinkesjatengprov.go.id
[4] Semarang city health office, 2015. Health Profile of Semarang city in 2015. Semarang, 2015, p.104. Available at: http://www.dinkes.semarangkota.go.id.
[5] U. F. Achmadi, *Region-Based Disease Management*, Jakarta, UI Press, 2008.
[6] B. Santosa, *Data Mining*, Yogyakarta, Graha ilmu., 2007
[7] J. Han, M. Kamber and J. Pei, *Data Mining: Concepts and Techniques*, 2011. Available at: http://link.springer.com/10.1007/978-3-642-19721-5.
[8] T. Santhanam, & M.S. Padmavathi,. Application of K-Means and genetic algorithms for dimension reduction by integrating SVM for diabetes diagnosis. 2014. Procedia Computer Science, 47(C), pp.76–83. Available at: http://dx.doi.org/10.1016/j.procs.2015.03.185.
[9] R.M.S.A. Ni'mah,. Geographic Information System Visualization of Clustering of ARI in Kaliwungu District , 2012, pp.5–6.
[10] S. Garg, & N. Rupal, A Data Mining Approach to Detect Tuberculosis Using Clustering and GA-NN Techniques. 2012, 4(10), pp.2013–2016.
[11] E.C. Fradelos, et al., Health based geographic information systems (GIS) and their applications. Acta Informatica Medica, 2014. 22(6), pp.402–405.
[12] E. Firmansyah, Geographic Information System: Basic Principles and Application Development I., Yogyakarta: Digibooks,2016
[13] D. Wardani, et al., Clustered tuberculosis incidence in Bandar Lampung , Indonesia. WHO South-East Asia Journal of Public Health, 2016/(3 June 2014), pp.5–7.
[14] S. Hassarangsee, N.K. Tripathi, & M. Souris,. Spatial pattern detection of tuberculosis: A case study of si sa ket province, Thailand. International Journal of Environmental Research and Public Health, 12(12), pp.16005–16018. 2015. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84950239229&partnerID=40&md5=0cf1581816e69c727328bf48ce1d5bb8.
[15] W. Sun, et al., A spatial, social and environmental study of tuberculosis in China using statistical and GIS technology. International Journal of Environmental Research and Public Health, 12(2), 2015, pp.1425–1448
[16] Gorunescu, Data Mining Concepts,Models and Techniques, Springer-Verlag Berlin Heidelberg (Series Intelligent Systems Reference Library), 2011.
[17] TS Venâncio, TS Tuan and LFC Nascimento, Indidence Of Tuberculosis In Children In The State Of São Paulo, Brazil, Under Spatial Approach, 2015, pp:1541-1546, available at : https://www.ncbi.nlm.nih.gov/pubmed/26017955
[18] K Middelkoop , LG Bekker, C Morrow, E Zwane and R Wood, Childhood tuberculosis infection and disease: A spatial and temporal transmission analysis in a South African township, S Afr Med Journal, 2009, pp: 738-743.

[19] D. Pfeiffer, T. Robinson, M. Stevenson, K Stevens, D. Rogers , A. Clements, A. Spatial Analysis in epidemiology. New York:Oxford University Press Inc.; 2008.
[20] C.E. Sabel, and S. Löytönen. Clustering of Disease: Disease Mapping and Spatial Analysis. GIS Public Heal. Pract. USA: CRC Press LLC; 2004. p. 142.

Clustering tuberculosis in children

ORIGINALITY REPORT

11%	7%	6%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	repository.ubaya.ac.id Internet Source	2%
2	www.ncbi.nlm.nih.gov Internet Source	1%
3	mpra.ub.uni-muenchen.de Internet Source	1%
4	www.mdpi.com Internet Source	1%
5	Submitted to Institut Pertanian Bogor Student Paper	1%
6	Siriwan Hassarangsee, Nitin Tripathi, Marc Souris. "Spatial Pattern Detection of Tuberculosis: A Case Study of Si Sa Ket Province, Thailand", International Journal of Environmental Research and Public Health, 2015 Publication	1%
7	Dewi Handayani, Shofi Nur Inayati, Amirotul Musthofiah Hidayah Mahmudah. "Modal choice	1%

between bicycle and pedicab using stated preference method in Benteng Vastenburg and Keraton Surakarta", AIP Publishing, 2019
Publication

8	Submitted to Postgraduate Institute of Medicine Student Paper	<1%
9	repository.maranatha.edu Internet Source	<1%
10	Submitted to Higher Education Commission Pakistan Student Paper	<1%
11	Submitted to University of Edinburgh Student Paper	<1%
12	eprints.walisongo.ac.id Internet Source	<1%
13	Submitted to University of Malaya Student Paper	<1%
14	Sandeep Kaur, Sheetal Kalra. "Disease prediction using hybrid K-means and support vector machine", 2016 1st India International Conference on Information Processing (IICIP), 2016 Publication	<1%
15	Submitted to University of Surrey Roehampton Student Paper	<1%

16	Submitted to University of South Florida Student Paper	<1%
17	G Lucarelli, A Isgrò, P Sodani, M Marziali et al. "Hematopoietic SCT for the Black African and non-Black African variants of sickle cell anemia", Bone Marrow Transplantation, 2014 Publication	<1%

Exclude quotes	Off	Exclude matches	< 5 words
Exclude bibliography	On		