Dengue Fever Outbreak Prediction in Surabaya using A Geographically Weighted Regression

by Tanti Octavia

Dengue Fever Outbreak Prediction in Surabaya using A Geographically Weighted Regression

Siana Halim
Industrial Engineering Department
Petra Christian University
Surabaya, Indonesia
halim@petra.ac.id

Felecia
Industrial Engineering Department
Petra Christian University
Surabaya, Indonesia
felecia@petra.ac.id

Tanti Octavia
Industrial Engineering Department
Petra Christian University
Surabaya, Indonesia
tanti@petra.ac.id

Andreas Handojo
Informatics Department
Petra Christian University
Surabaya, Indonesia
handojo@petra.ac.id

***Abstract*— Dengue Fever is one of the viral diseases of the tropics that are easily spread in high density and humid area such as in Surabaya. Many researchers in various expertise have studied this disease. Some of them use statistical and machine learning approach to predict the outbreak of the disease, so that the government can prevent that incident. In this paper we use the geographically weighted regression for predicting the dengue fever outbreak in Surabaya. The geographically weighted regression has superiority in estimating the coefficient of the explanatory variables locally. So that, we can put more attention into the region with has high estimates coefficient parameters. Here, we look at the locally estimates of the dengue fever infected in the year 2016, 2017, population density and poverty percentage for predicting the dengue fever outbreak in the year 2018. In this study, the pattern of the predicted model can follow the pattern of the true dataset.**

Keywords*— *Dengue fever outbreak; Global Moran I statistics; Local Moran I statistics; Geographically Weighted Regression; locally parameters estimate

I. INTRODUCTION

Surabaya with its 3.5 million populations (estimated) is the second largest city in Indonesia and also the capital of East Java Province [1]. Surabaya play important economic role in Indonesia as a port city since it was established in 1293. Many people from other cities and country come to Surabaya to do business. This has caused city density reach 9900 peoples /km2 [2] (higher than Singapore 8108/km2 or Hong Kong 6677/km2). A high-density city like Surabaya has many socioeconomic problems for its citizens, and it is closely related to its citizen health and wellbeing.

Dengue Fever is one of the viral diseases of the tropics that are easily spread in high density and humid area like Surabaya [3, 4]. Dengue Fever usually spread during wet season in tropical country around month December with its peak around March and April. There is 2660 Dengue Fever case in East Java Province in January 2019, and 46 victims are died [5]. There are 42 Dengue Fever victims in Surabaya in 2018 and one victim died. This year (2019) Surabaya city government has managed to control the Dengue Fever spread so only 23 victims reported [6].

Many researches have been done to predict the dengue fever in several countries. Using artificial neural network, [7] predicted the Dengue Hemorrhagic fever outbreak in Thailand, [8] predicted the Dengue outbreak in Srilanka, [9] predicted Dengue Fever outbreak in the Northwest Coast of Yucatan, Mexico and San Juan, Puerto Rico. Hartanto et al. [10], applied the geostatistics for predicting the Pneumonia patients in Surabaya-Indonesia. In Sleman, a district in Central Java Indonesia [11] predicted dengue hemorrhagic fever (DHF) using vector autoregressive spatial autocorrelation (varsa). Mahdiana et al. [11] used 5 years dataset to predict the DHF outbreaks. In the model, they include min, max and average temperature, average humidity, and rainfall and irradiation time. Instead of using neural network approach, classical geostatistics and varsa, in this study we use the geographically weighted regression (GWR). The spatial autoregressiove models [12] have assumption that the structure of the models remains constant, i.e., there is no local variations in the parameter estimates. The GWR [13] allows the estimated parameters vary locally. Therefore, to explore the dengue fever outbreak in Surabaya, we will use the GWR.

II. METHODS

In this study, we first did the data descriptive of the Surabaya population density, the area, percentage of poverty, the average temperature and humidity in each district. We then defined the polygon of each district in which the community health centers (puskesmas) are located. Two polygons P_i and P_j are neighbors if they share a common boundary. To

We tested the spatial global autocorrelation of the deque fever using Moran I statistics [14]. The Global Moran's I statistic, has the null hypothesis that states the observed random variable is randomly distributed against it has spatial pattern. The Global Moran's I statistics can be formulated as:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (1)$$

We also tested the spatial local autocorrelation of the deque fever using local Moran I statistics. The local Moran's I statistics was suggested by Anselin [15] for identifying local clusters and local spatial outliers. The local Moran's I statistics can be used to classify the significant locations as high-high and low-low spatial clusters, and high-low and low-high spatial outliers. Many researchers have applied the local Moran's I statistics for their study, e.g. [16] and [17]. The Local Moran's I statistics can be formulated as:

$$I_i = \frac{n(y_i - \bar{y})\sum_{j=1}^{n} w_{ij}(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (2)$$

where y is the variable of interest.

After testing the spatial local autocorrelation, we then modelled the dengue fever in Surabaya using the geographically weighted regression [13]. The geographically weighted regression models [13] can be formulated as

$$y_i = X\beta_i + \varepsilon \qquad (3)$$

where i is the location in which the local parameters will be estimated.

The β_i is the parameters at the location i and can be estimated as

$$\beta_i = (X'W_iX)^{-1}X'W_iy \qquad (4)$$

where w_{ij} is the weight for the j observation and formulated as the Gaussian function

$$w_{ij} = e^{\left(\frac{-d_{ij}}{h}\right)^2} \qquad (5)$$

The d_{ij} is the Euclidean distance between the location of observation i and location j, while h is the bandwidth. The bandwidth h can be selected such that the root mean square prediction error is minimum.

III. Result and Discussion

The data were collected data from 63 community health centers (pusat kesehatan masyarakat) in Surabaya. The outbreak was worse in 2016, but it is under controlled so that the number of the dengue fever infected is decreased significantly (Fig. 1). The Government of Surabaya has a program so called "Jumantik-Juru Pemantau Jentik", they are usually housewives who have a duty to observe the mosquito larva in houses and environment in their area. They sampled 20 houses in each area and recorded either there is mosquito larva or not in the sampled house. If the mosquito larva is founded in a house then they will calculate the number of free mosquito larva as (19/20) * 100%, that is 95% of the houses in that area is free from mosquito larva. Jumantik will give reports about the number of mosquito larva found and number of people infected to dengue fever every month to health center in the area (Fig. 2). Surabaya government focus more on preventive action to reduce dengue fever outbreak. Therefore, health center will promote dengue prevention through environmental cleaning programs especially during wet season [18].

The effort of the Surabaya government to reduce the dengue fever can be seen in Fig. 3. It shows the linear trend the % mosquito larva free vs the number of dengue fever infected in 2016 dan 2017. The trend in 2016 is increasing, it means that the percentage of mosquito larva free is increasing the number of dengue fever infected is also increasing. In the opposite, the trend in 2017 is decreasing. The larger the percentage of mosquito larva free, the smaller the number of dengue fever infected. Those two trends indicate that the preventive action of Surabaya government is successful.

However, both trends have small r-squared statistics, it indicates that the correlation is weak. Therefore, in this paper we will study the spatial effect of the dengue fever outbreak in Surabaya. We collected data of number if rainy days in a year, precipitation, maximum and minimum temperature, maximum and minimum humidity, population density, and poverty percentage in each Surabaya's district [19]. The summary statistics for the Surabaya in 2018 can be seen in Table 1.

TABLE I. Surabaya Statistics in 2018

	Min	Mean	Max
Population (thousand)	12541	45802	87561
Area (Km2)	0.915	2.001	14.400
Density (thousand/Km2)	2733	46992	541022
Poverty percentage (%)	4.03	18.02	55.46
#Rainy day (days/month)	9.83	13.99	16.00
Precipitation (mm/month)	129.9	164.6	194.9
Max Humidity per month	70	88.72	94.75
Min Humidity per month	46.08	53.14	57.83
Max Temperature	28.21	33.30	34.43
Min Temperature	23.11	26.29	28.73

At the first step we use the global Moran's I statistics to test the data are under randomization (Ho) or have spatial dependencies. The test shows that the data is significantly have spatial dependencies (p-value = 0.0154 (Fig. 4). To see which districts is spatially correlated strongly we then use the local Moran's I statistics. Fig. 5 shows that there are two districts which have very strong spatial correlation, they are Dukuh Kupang and Putat Jaya, and both districts have 13 and 17 dengue fever infected consecutively in 2018.

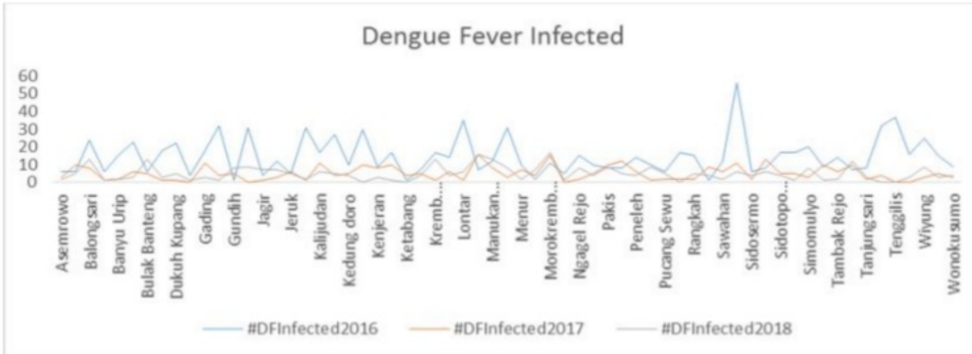


Fig. 1. The numbers of Dengue Fever Infected from 2016-2017 in each community health centers.

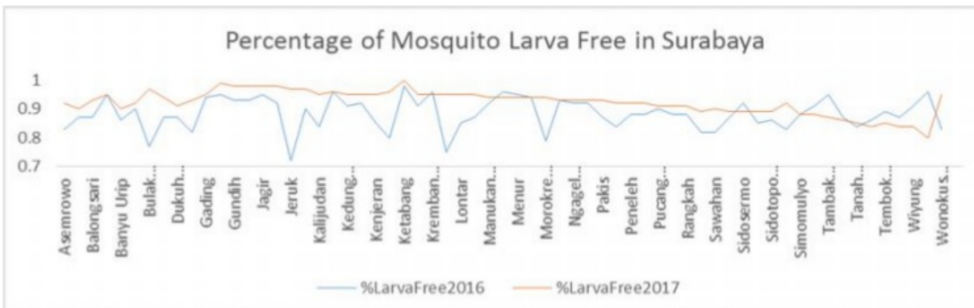


Fig. 2. Percentage of mosquito larva free in each Surabaya's districts

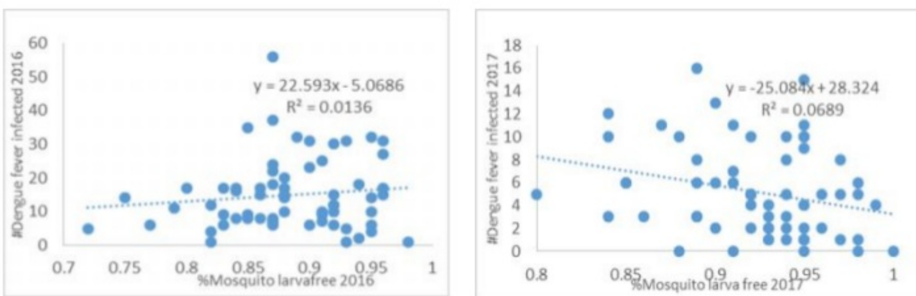


Fig. 3. Linear trend the % mosquito larva free vs the number of dengue fever infected in 2016 dan 2017

Moran I test under randomisation

data: data$DB2018

weights: W_cont_el_mat n reduced by no-neighbour observations

Moran I statistic standard deviate = 2.3064, p-value = 0.01054

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
0.27393001	-0.01923077	0.01615629

Fig. 4. Global Moran I statistics for dengue fever infected in the year 2018

The geographically weighted regression (GWR) models permit the parameters estimate vary locally in each district in which the community health centers locate. Table 2 presents the summary of GWR coefficient estimates at data points. The number of rainy days, precipitation, max and min humidity, and temperature are not varied too much since those community health centers are locate in the same climate. Therefore, in we only look at the local coefficient estimates at the dengue fever infected (DFI) 2016 and 2017, population density and poverty percentage.

We divide the coefficients estimates at data points into five intervals. Those intervals represent very low, low, medium, high and very high effect for each explanatory variable in each district (see Fig. 6). The DFI in 2016 in Keputih (K7), Medokan Ayu (M3) and Gunung Anyar (G4) have strongest correlation to the DFI 2018, while Benowo (B5), Sememi (S3), Made (M1) and Jeruk (J3) have the weakest correlation to the DFI 2018. Similarly, we can see in Fig. 6, the strongest and the weakest correlation for the DFI 2017, population density and poverty percentage to the DFI 2018.

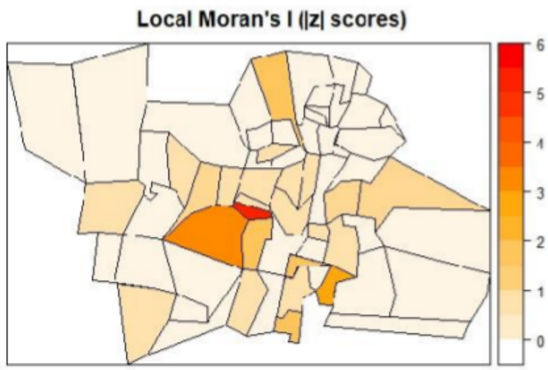


Fig. 5. Local Moran I statistics for the dengue fever infected in the year 2018.

TABLE II. SUMMARY OF GWR COEFFICIENT ESTIMATES AT DATA POINTS

	Global	Min	Median	Max
Intercept	-3.688	-5.522	-4.237	-3.386
RainyDays	-0.211	-0.236	-0.213	-0.180
Percipitation	0.015	0.014	0.015	0.018
MaxHumidity	-0.130	-0.136	- 0.127	-0.121
MinHumidity	-0.039	-0.043	-0.028	-0.015
MaxTemp	0.582	0.552	0.574	0.609
MinTemp	-0.055	-0.059	-0.054	-0.046
Pop_density_Scale	2.537	2.077	2.446	2.783
%Povery	0.008	0.006	0.007	0.010
DFI2016	0.106	0.093	0.111	0.124
DFI2017	0.397	0.365	0.394	0.427

Finally, using the GWR model we plot the dengue fever infected in 2018 vs the predicted (Fig. 7). The predicted can follow the pattern of the true dataset, the MSE of this prediction is 8. 59, this due to in some locations, e.g. Pegirian, Pucang Sewu, the errors prediction is high.

IV. CONCLUSION

We modeled the dengue fever infected in Surabaya using geographically weighted regression. The GWR not only produce a model, but also can show the effect of each coefficients estimates for each explanatory variable locally. For the moment, the MSE of this prediction is still high due to in some locations, e.g. Pegirian, Pucang Sewu, the errors prediction is high. This problem leads us to the future research for finding a good model to represent the dengue fever infected prediction with small mean square error.

ACKNOWLEDGMENT

The authors are very grateful to the Directorate of Higher Education of The Republic of Indonesia for financial support of this research. We also thank to the Surabaya Public Health Office for providing the data.

REFERENCES

[1] Statistics Indonesia. "Surabaya Municipality in Figures 2017". surabayakota.bps.go.id. Archived from the original on 2019-04-01. Retrieved 2019-04-01.

[2] "Indonesia: Java (Regencies, Cities and Districts) – Population Statistics, Charts and Map".

[3] Kularatne SA (September 2015). "Dengue fever". BMJ. 351: h4661. doi:10.1136/bmj.h4661. PMID 26374064.

[4] "Dengue and severe dengue Fact sheet N°117". WHO. May 2015. Archived from the original on 2 September 2016. Retrieved 3 February 2016.

[5] Kasus demam berdarah di Jawa Timur, https://regional.kompas.com/read/2019/01/30/21522801/ada-2660-kasus-demam-berdarah-di-jatim-46-penderita-meninggal.

[6] https://kelanakota.suarasurabaya.net/news/2019/216123-Puncak-Wabah-DBD-Diperkirakan-April-dan-Maret,-Pemkot-Lakukan-Cegah-Dini

[7] B. Jongmuenwai, S. Lowanichchai, and S. Jabjone, "Prediction model of Dengue Hemorrhagic fever outbreak using artificial neural networks in Northeast of Thailand", International journal of pure and applied mathematics, vol. 118, no. 18, pp3407-3417, 2018.

[8] P.H.M.N. Herath, A.A.I. Perera, and H.P.Wijekoon, "Prediction of Dengue Outbreaks in Srilanka using Artificial Neural Network", International Journal of Computer Applications, vol.101, pp. 1-5, September 2014.

[9] A.E. Laureano-Rosario, A.P Duncan, P.A. Mendez-Lazaro, J.E. Garcia-Rejon, S. Gomez-Carro, J. Farfan-Ale, D.A. Savic, and F.E. Muller-Karger, "Application of artificial neural networks for Dengue Fever Outbreak Predictions in the Northwest Coast of Yucatan, Mexico and San Juan, Puerto Rico", Tropical Medicine Infectious Disease, vol 3, no. 5, pp. 1-16, 2018.

[10] S. Hartanto, S. Halim, and O.Y. Yuliana, "Pemetaan penderita pneumonia di Surabaya, dengan menggunakan geostatistika", Jurnal Teknik Industri, vol. 12, no.1, pp. 41-46, 2010.

[11] D. Mahdiana, E. Winarko, A. Ashari, and H. Kusnanto, "A model for forecasting the number of cases and distribution pattern of Dengue Hemorrhagic fever in Indonesia", International journal of advanced computer science and applications, vol. 8, no. 11, pp. 143-150, 2017.

[12] J.M. Ver Hoef, E.E. Peterson, M.B Hooten, E.M. Hanks, and M.J. Fortin, "Spatial autoregressive models for statistical inference from ecological data", Ecological society of America, vol. 88, issue 1, pp. 36-59, 2018.

[13] A.S. Fotheringham, C. Brunsdon, C., and M.E. Charlton, "Geographically weighted regression: The analysis of spatially varying relationships". Wiley, Chichester, 2002.

[14] L. Hongfei, C.A. Calder, N. Cressie, "Beyond Moran's I: Testing for spatial dependence based on the spatial autoregressive model", Geographical Analysis, vol. 39, no. 4.pp. 357-375, 2007.

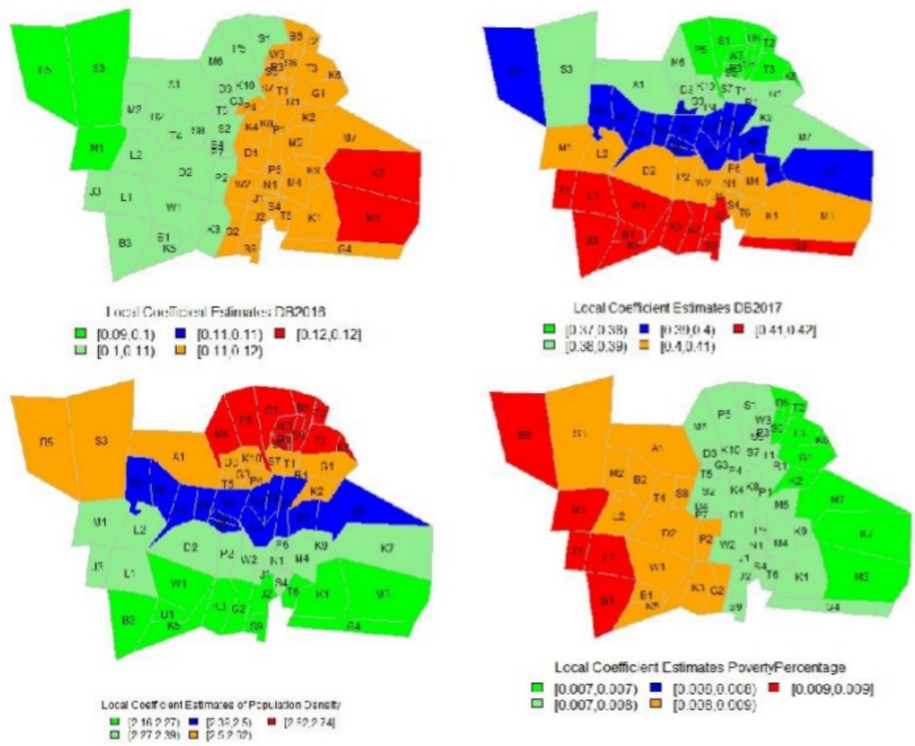
[15] L. Anselin, "Local indicators of spatial association-LISA", Geographical Analysis, vol. 27, pp. 93-115, 1995.

[16] C. Zhang, L. Luo, W. Xu, and V. Ledwith, "Use of local Moran 's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland", Science of the total environment, vol. 398, issue 1-3, pp. 212-221, 2008.

[17] Y. Yuan, M. Cave and C. Zhang, "Using local Moran's I to identify contamination hotspots of rare earth elements in urban soils of London", Applied geochemistry, vol. 88, part B., pp. 167-178, 2018.

[18] Jumantik, https://surabaya.kompas.com/read/2018/11/01/16393191/gerakan-1-rumah-1-jumantik-langkah-risma-agar-surabaya-bebas-dbd

[19] Surabaya dalam Angka, Biro Pusat Statistika Surabaya, https://surabayakota.bps.go.id/publication/2018/08/21/35de76f19338e3ecd445b838/kota-surabaya-dalam-angka-2018.html



A1	Asemrowo	B1	Balas Klumprik	B2	Balongsari	B3	Bangkingan	B4	Banyu Urip
B5	Benowo	B6	Bulak Banteng	D1	dr. Soetomo	D2	Dukuh Kupang	D3	Dupak
G1	Gading	G2	Gayungan	G3	Gundih	G4	Gunung Anyar	J1	Jagir
J2	Jemursari	J3	Jeruk	K1	Kali Rungkut	K2	Kalijudan	K3	Kebon Sari
K4	Kedung Doro	K5	Kedurus	K6	Kenjeran	K7	Keputih	K8	Ketabang
K9	Klampir Ngasem	K10	Krembangan Selatan	L1	Lidah Kulon	L2	Lontar	M1	Made
M2	Manukan Kulon	M3	Medokan Ayu	M4	Menur	M5	Mojo	M6	Morokrembangan
M7	Mulyorejo	N1	Ngagel Rejo	P1	Pacar Keling	P2	Pakis	P3	Pegirian
P4	Peneleh	P5	Perak Timur	P6	Pucang Sewu	P7	Putat Jaya	R1	Rangkah
S1	Sawah Pulo	S2	Sawahan	S3	Sememi	S4	Sidosermo	S5	Sidotopo
S6	Sidotopo Wetan	S7	Simolawang	S8	Simomulyo	S9	Siwalankerto	T1	Tambak Rejo
T2	Tambah Wedi	T3	Tanah Kali Kedinging	T4	Tanjungsari	T5	Tembok Dukuh	T6	Tenggilis
W1	Wiyung	W2	Wonokromo	W3	Wonokusumo				

Fig. 6. The local coefficient estimates of Dengue Fever Infected in 2016, 2017 (DB2016, DB2017), population density dan povery percentage

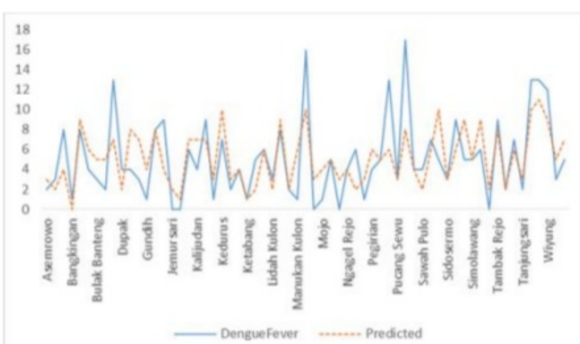


Fig. 7. The data vs the predicted dengue fever infected in 2018 using GWR

Dengue Fever Outbreak Prediction in Surabaya using A Geographically Weighted Regression

ORIGINALITY REPORT

10%	7%	8%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	file.scirp.org Internet Source	1%
2	ijarai.thesai.org Internet Source	1%
3	K.G.S. Dharmawardana, J.N. Lokuge, P.S.B. Dassanayake, M.L. Sirisena, M.L. Fernando, A.S. Perera, S. Lokanathan. "Predictive model for the dengue incidences in Sri Lanka using mobile network big data", 2017 IEEE International Conference on Industrial and Information Systems (ICIIS), 2017 Publication	1%
4	Apostolos Ziakopoulos, George Yannis. "A review of spatial approaches in road safety", Accident Analysis & Prevention, 2020 Publication	1%
5	Kai Fang, Tingting Wang, Jianjian He, Tijian Wang, Xiaodong Xie, Yiqi Tang, Yang Shen, Anqi Xu. "The distribution and drivers of PM2.5	1%

in a rapidly urbanizing region: The Belt and Road Initiative in focus", Science of The Total Environment, 2020
Publication

6	Gamerman, D.. "Multivariate spatial regression models", Journal of Multivariate Analysis, 200411 Publication	1%
7	pcneedtogo.blogspot.com Internet Source	1%
8	Vicka Kharisma, Naoya Abe. "Food Insecurity and Associated Socioeconomic Factors: Application of Rasch and Binary Logistic Models with Household Survey Data in Three Megacities in Indonesia", Social Indicators Research, 2019 Publication	1%
9	open.uct.ac.za Internet Source	1%
10	Submitted to Higher Education Commission Pakistan Student Paper	1%
11	Submitted to University of Malaya Student Paper	1%

Exclude quotes	Off	Exclude matches	< 1%
Exclude bibliography	Off		