Malaysian Journal of Mathematical Sciences

Journal homepage: https://einspem.upm.edu.my/journal

Grouping District and Cities for Environmental Sanitation Indicators of SDGs Achievement in East Java

Permatasari, E. O* and Zain, I.

Department of Statistics, Institut Teknologi Sepuluh Nopember, Indonesia

E-mail: erma.oktania@gmail.com
**Corresponding author*

Received: 26 April 2019
Accepted: 2 June 2021

Abstract

Environmental sanitation is the health status of an environment that includes housing, sewerage, and clean water supply. Guaranteeing the availability and management of sustainable water and sanitation are the global goals of Presidential Regulation Number 59 of 2017 signed by President Joko Widodo on July 4, 2017, to fulfill the government's commitment in achieving the achievement of Sustainable Development Goals (SDGs). East Java Province, as one of the provinces that do Presidential Regulation Number 59 of 2017, make efforts to improve the condition of environmental health and basic sanitation by implementing Total Sanitation Based Community (STBM). STBM is an activity that focuses on integrated preventive and promotive efforts to trigger and maintain the sustainability of changes in the behavior of clean and healthy living communities. Despite efforts to improve environmental health and basic sanitation conditions have been done, but there are still gaps in some district and cities in East Java Province. Suppose for a healthy house, the coverage for Trenggalek District that fulfill for healthy house is 93.97%, while for Sumenep District is 6.69%. The percentage of district and cities that have access to proper sanitation, with the best sanitation access is Madiun City of 100% and the lowest is Situbondo District of 13%. Based on the problem, this research was done for grouping district and cities in East Java Province based on environmental sanitation indicators, so it is expected to provide information about the groups of districts and cities that have good access and the less good of environmental sanitation. This information can also be a reference for East Java Provincial Health Office in making policies or improvement efforts for regions belonging to district and cities that have poor environmental sanitation access. The grouping was done using robust clustering using links (ROCK) method. The results show that the optimum grouping was two groups.

Keywords: sanitation; district; cities; grouping; East Java province.

1 Introduction

Environmental sanitation is the health status of an environment. It includes housing, sewage disposal and clean water supply. Environmental sanitation is intended to meet the requirements of a healthy and comfortable environment. Some indicators that describe environmental conditions include healthy houses, clean water and basic sanitation facilities such as waste water disposal, trash bins, and ownership of latrines and waste treatment facilities at health care facilities [1]. The guarantee of the availability and sustainable management of clean water and sanitation is also one of the global goals of Presidential Regulation Number 59 of 2017 signed by President Joko Widodo on July 4, 2017, to fulfill the government's commitment to the achievement of the Sustainable Development Goals (SDGs).

East Java Province, as one of the provinces running Presidential Regulation Number 59 of 2017, made efforts to improve environmental health conditions and basic sanitation by carrying out Community Based Total Sanitation (STBM) activities. STBM is an activity that focuses on integrated preventive and promotive efforts through efforts to trigger and maintain the sustainability of changes in people's behavior in clean and healthy living, so that the community is expected to play an active role in the provision of basic sanitation through three STBM components which consist of efforts to create needs (demand), service provision (supply) and the creation of a enabling environment (enabling environtment). The three STBM components form are the basis of the implementation strategy for the achievement of the five STBM pillars, namely Open Defecation (SBS), Hand Washing with Soap (CTPS), drinking water and household food processing (PAMM-RT), Household Waste Management (PS-RT), and Household Liquid Waste Processing (PLC-RT).

Although efforts to improve environmental health conditions and basic sanitation have been carried out, there are still gaps in several district and cities in East Java Province. For example, for a healthy house, the coverage for Trenggalek District that meets healthy homes is 93.97%, while for Sumenep District is 6.69%. The percentage of district and cities that have access to proper sanitation (people who access communal latrines and goose-necked toilets and are contemplated plus recharge), with the best access to proper sanitation is 100% of Madiun City and the lowest is Situbondo District at 13%. If sorted according to proper sanitation, the use of goose-necked latrines and septictank in East Java Province as a whole is 66.96%. Poor sanitation also causes various diseases, one of which is diarrheal disease. Based on the health profile of East Java Province, from 2009 to 2014, the achievement of diarrhea case findings tends to increase every year and needs to be aware of the increase in cases in February, July, and November. While the peak of cases occurred at the beginning of the year, namely January.

Based on this background, this research will carry out grouping district and cities in East Java Province based on indicators of environmental sanitation, so that it is expected to provide information about groups of districts and cities that have good and bad access to environmental sanitation. This information can also be an information for the East Java Provincial Health Office in making policies and to improve efforts for districts and cities that have poor access to environmental sanitation. Research on this matter has never existed before, especially by using data measurement which is entirely qualitative data.

The problem raised in this research is how to group districts and cities in East Java Province based on environmental sanitation indicators. Then make the descriptive statistics are used to describe the results of the grouping obtained along. The results of this research can contribute to the development of science and technology, especially for developing clustering method in categorical data. This research is also very useful for interested parties, especially the East Java Provincial

Health Office in classifying districts and cities in East Java Province based on indicators of environmental sanitation as a basis for evaluating the achievement of SDGs goals.

2 Related Works

Several studies on sanitation have been conducted. For example, [7] determined the relationship of environmental sanitation conditions with the incidence of diarrhea in children under five in the work area of Baranti Health Center Sidrap District in 2013. Then [10] examined the relationship of knowledge, behavior, and environmental sanitation with numbers helminthiasis for elementary school children in Palu City. Furthermore [4] examined the relationship between the condition of home basic sanitation and the incidence of diarrhea in children under five in the working area of the Rembang Health Center 2. There [8] who examined the relationship between the quality of environmental and bacteriological sanitation of clean water to the incidence of diarrhea in children under five in the Adiwerna Health Center in Tegal District.

While research on grouping analysis was conducted by [5] who used the Multivariate Adaptive Regression Spline (MARS) approach to classification of STI-infected status in CSWs in Moroseneng Surabaya localization. Then In addition, [6] also once categorized the interest in the ITS SBMPTN.

3 Materials and Methods

3.1 Basic Environmental Health and Sanitation Services

To minimize the risk of disease or health problems as a result of an unhealthy environment, various efforts have been made to improve the quality of the environment. Some indicators that describe environmental conditions include healthy houses, clean water and basic sanitation facilities such as waste water disposal, trash bins and ownership of latrines and waste treatment facilities in health care facilities. In an effort to improve the conditions of environmental sanitation and basic sanitation in East Java, Community Based Total Sanitation (STBM) activities have been carried out. STBM is an activity that focuses on integrated preventive and promotive efforts through efforts to trigger and maintain the sustainability of changes in people's behavior in living clean and healthy, so that the active role of the community in the provision of basic sanitation through three components of STBM consists of: efforts to create needs (demand), service provision (supply) and the creation of a conducive environment (enabling environment).

These form the basis of implementing an adequate strategy to achieve the five STBM pillars which are:

i. Stop Open Defecation (SBS).
ii. Handwashing with soap (CTPS).
iii. Drinking water treatment and household food (PAMM-RT).
iv. Household Waste Management (PS-RT).
v. Household Liquid Waste Management (PLC-RT).

3.1.1 Healthy Home

Healthy Houses are residential buildings that meet health requirements which consist of home components, sanitation facilities and behavior, among others, namely having healthy latrines, landfills, clean water facilities, suggestions for waste water disposal, good ventilation, suitable housing density and house floors. from the ground. In 2015 the number of healthy houses in East Java was 5,482,738 (54.60%), and in 2016 coaching was carried out by 2,056,633 and those who fulfilled the requirements of 1,185,086 (57.62%) houses. So that in 2016 there were 6,667,824 (63.34%) of all houses in East Java. The highest coverage of healthy homes is Trenggalek District with 93.97% coverage. While the lowest coverage is occupied by Sumenep District with a coverage of 6.69%. However, overall each district and cities experienced an increase, with the highest increase in Batu City by 37.74% and the lowest Kediri Citywhich increased 0.1%.

Guidance and supervision are the funtions of Health Service, so that the improvement of healthy house is from a monitoring point of view and not development. To increase the coverage of healthy homes in East Java, each district and cities has several activities including community empowerment using participatory methods and TMND (TNI Manunggal Enter Village). The method is by giving stimulants which last year were still given to under privileged people and high risk diseases of environmental based.

3.1.2 Organizing Drinking Water

Drinking water sampling is carried out based on the results of sanitation inspection, namely on drinking water with piping systems, drinking water depots and drinking water instead of piping networks with moderate and low pollution risk. The frequency of sanitation inspection is carried out during the dry season and the rainy season. Drinking water operators are state-owned or regional-owned business entities, cooperatives, private business entities, individual businesses, community either groups or individuals who carry out the provision of drinking water supply.

The results of the implementation of drinking water, in East Java Province in 2016 there were 5,307 drinking water providers and there are 7,085 water samples have been examined. Examinations carried out include physical, bacteriological and chemical examinations. Of the 7,085 samples examined, 5,458 samples or 77.04% had met the requirements, there was a quality improvement of 2.67% compared to 2015 with only 74.37% of the results of drinking water administration, in East Java Province in 2016 there were 5,307 organizers drinking water. And a sample of 7,085 samples have been examined. Examinations carried out include physical, bacteriological and chemical examinations. Of the 7,085 samples examined, 5,458 samples or 77.04% had met the requirements, there were 2.67% quality improvements compared to 2015 which only 74.37%.

3.1.3 Community-Based Total Sanitation (STBM)

Sanitation development in Indonesia has shown significant progress. This can be seen from the achievement of healthy latrine access in 2016 of 67%. However, this achievement is the total community that accesses both the goose neck, cemplung and plengsengan. If it is sorted according to proper sanitation, the goose-necked latrines are 66.96%. This achievement was motivated by the establishment of Community-Based Total Sanitation (STBM) as a national sanitation development strategy in 2008, which was later updated and strengthened by the Republic of Indonesia's Minister of Health Regulation No. 3 of 2014 concerning STBM.

Please note there are two understandings of proper sanitation and healthy latrines. What is meant by proper sanitation here is the community that accesses communal latrines and goose-necked toilets and is contradictory plus recharge. Whereas the latrines are healthy in addition to the above two as well as plunge pit latrines. These two variables are included in healthy latrines because there are certain areas that are indeed difficult to get clean water, so it is not possible to build a goose-necked latrine.

In East Java, there are 4 districts and cities whose communities have not defecated anywhere or open places, namely Pacitan, Magetan, Ngawi and Madiun. Of the four cities, Madiun, which has access to sanitation, is as good as 100% and the lowest is Situbondo. Proper sanitation is also influenced by the level of education and the increasingly advanced culture of society.

3.1.4 Clean and Healthy Life Behavior

Percentage of households with Clean and Healthy Life Behavior (PHBS) was obtained by using the ten indicators stated below:

i. Delivery assistance by health workers.
ii. Babies given exclusive breastfeeding Babies are given exclusive breastfeeding.
iii. Toddlers are weighed every month.
iv. Use clean water.
v. Wash hands with clean water and soap.
vi. Using healthy latrines.
vii. Eradicate larvae at home once a week.
viii. Eat vegetables and fruit every day.
ix. Do physical activity every day.
x. Don't smoke inside the house.

In a survey conducted in 2016, households with PHBS were found to be 53.82% and this was an increase of 1.97% from 51.85% observed in 2015. It was discovered that smoking in the house and exclusive breastfeeding were prioritized.

3.2 Grouping Techniques on Categorical Data

Traditional grouping algorithms using the concept of distance between objects to classify are not suitable for categorical data. A robust grouping hierarchy of the ROCK (Robust Clustering Using Links) algorithm is developed that uses links to group clusters [9]. Grouping based only on proximity between objects, is not strong enough to differentiate the two so that between groups is inseparable or well differentiated because it is quite possible for objects from different groups to neighbor. The link concept based approach adopts a comprehensive approach to grouping problems, where this concept can capture the overall knowledge of neighboring data objects into

relationships between individual pairs, and this concept is used by the ROCK method. The Robust method of ROCK is because this method uses the concept of links and not distance to group objects or clusters [3].

Links are used to measure similarity or distance between pairs of data points [3]. Easily, the number of links between two objects is the number of common neighbors they have from the initial data. After calculating the number of links between data objects, the algorithm starts with each cluster into one object and remains clustered into clusters based on the size of goodness for clustering. Observations that have a high level of link are combined into one group, while those with a low level of link are separated from the group where the observations are grouped.

The ROCK method can handle outliers quite effectively [3]. The pruning of the outliers allows the disposal of none of the neighbors, in order to avoid their participation in the grouping. But in some situations, they can be present in small groups. Therefore, the steps involved in the category data through the use of the ROCK method are as follows [2]:

i. Initialize each object as a cluster at the beginning of the process.

ii. Calculating Similarity. Similarity between two objects can be defined by some precise guesses of the distance between data objects. The size of the resemblance between the pairs of objects i and the object j can be calculated like equation (1):

$$sim(\,T_i, T_j) \quad = \quad \frac{|T_i \cap T_j|}{|T_i \cup T_j|}, i \neq j \tag{1}$$

where

$i = 1, 2, 3, \ldots, n; \quad j = 1, 2, 3, \ldots, n;$
$T_i = \{x_{1i}, x_{2i}, x_{3i}, \ldots, x_{m_{kategori}i}\}; T_j = \{x_{1j}, x_{2j}, x_{3j}, \ldots, x_{m_{kategori}i}\}.$

iii. Determine Neighbors. Pairing between objects and defined neighbors if the value, where the threshold value ranges from 0 to 1. The threshold value is determined by the researcher, depending on how large two objects or clusters can be said to be neighbors by the researcher, because the greater the threshold value will result in more objects or cluster that can be said to be neighboring.

iv. Calculate Links. Links between objects are defined as the number of common neighbors between and. If the link is bigger then most likely and are in the same cluster. The link matrix is obtained from the neighboring matrix multiplication with the neighboring matrix itself.

v. Determine local heap. Local heap is the value of goodness measure for each group with other groups if the link is $\neq 0$. Goodness Measure is an equation that calculates the number of links divided by the possibility of links formed based on the size of the group. Goodness Measure is obtained using the equation (2).

$$g(\,C_i, C_j) \quad = \quad \frac{link[\,C_i, C_j]}{(\,n_i + n_j^{1+2f(\,\theta)}\,) \; - n_i^{1+2f(\,\theta)} \; - n_j^{1+2f(\,\theta)}}, \tag{2}$$

where, $link[\,C_i, C_j] \; = (\,T_i, T_j)$, n_i and n_j are amount members of the group-i and group-j, while $f(\,\theta) \; = \frac{1-(\,\theta)}{1+(\,\theta)}$.

vi. Determine global heap. Global heap is the maximum value of goodness measure between columns in row i.

vii. Repeat steps (5) and (6) to get the maximum value in the global heap and local heap.

viii. During the data size > k, with k is the number of classes specified. So, a group combination has the largest local heap value with the largest global heap into one group. Then add links between groups that are merged, delete groups that are merged from the local heap and update the global value of the heap with the merged results.

ix. Perform step (8) to find the number of groups that are expected (k) or groups are raised automatically when there are no more links between groups. Grouping will continue to be conducted and stop when [[2]]:

   (a) The number of groups needed or desired has been met, or

   (b) There are no links between clusters.

3.3 Data and Variables

The data used in this research is secondary data obtained from the East Java Provincial Health Office and the Central Statistics Agency (BPS) in 2016. Several variables (independent variables $= x$) included in this research are shown in Table 1.

Table 1: Research variables.

Variable	Description	Data Type	Coding Category
x_1	The use of PAM water	Categorical	0.PAM 1.Don't PAM
x_2	The main drinking water	Categorical	0.Buy 1.Don't Buy
x_3	The main cooking water	Categorical	0.Buy 1.Don't Buy
x_4	The main bath / washing water	Categorical	0.Buy 1.Don't Buy
x_5	Distance of Drinking Water Source with Waste Shelter	Categorical	0.$<$ 10 meter 1.$\geq$ 10 meter
x_6	Distance of Cooked Water Source with Waste Shelter	Categorical	0.$<$ 10 meter 1.$\geq$ 10 meter
x_6	Distance of Bath Water Source with Waste Shelter	Categorical	0.$<$ 10 meter 1.$\geq$ 10 meter

3.4 Data Analysis Method

The steps involved in the analysis are as follows:

. Collect sample data from the East Java Provincial Health Office and the Central Statistics Agency (BPS).

. Research literature through books, journals, and field observations.

. Make a description of the characteristics of each variable in the district and city of East Java Province.

. Establish grouping of districts and cities in East Java Province based on environmental sanitation indicators by Robust Clustering using linKs (ROCK). The aim of this research is to carry out grouping districts and cities in East Java Province based on indicators of environmental sanitation, and all of these indicators are categorical types. ROCK method is used because it has better categorical data handling. This method also uses important concepts in grouping, namely defining links between neighbors, so that they can produce more valid groupings [[9]].

. Make descriptive statistics to describe the results of the groupings obtained. Descriptive statistics are used to describe the condition of each group formed based on the independent variables used in this study. At this stage there are a list of districts and cities in East Java Province that are still low at indicators of environmental sanitation, especially regarding clean water and good sanitation, as a basis for evaluating achievement of objectives SDGs.

4 Results

The data can be analyzed by characteristic of categorical data and grouping categorical data. The following is the discussion.

4.1 Characteristics of Categorical Data

Figure 1 shows that the main drinking water source that is most widely used in East Java is from the borehole or pump which is 24.55%, while the smallest is from other sources that is equal to 0.05%. Next to Figure 2 shows the percentage of the main drinking water sources that are distinguished between buying or not buying categories. The main drinking water sources that are categorized as buying are sources of drinking water obtained from branded bottled water, refill water, meter plumbing, and retail plumbing. Meanwhile the main drinking water sources categorized as non-buying are sources of drinking water obtained from bore wells or pumps, protected and unprotected wells, protected and unprotected springs, surface water, rainwater, and other sources. Based on these 2 new categories (buy and not buy), there are 64% of all RTs in East Java who get their drinking water source mainly from not buying, while the remaining 36% get their main drinking water source by buying.

Types of latrines that are widely used include goose neck, plengsengan close, plengsengan without lid, and cemplung or cubluk. Figure 3 shows the most widely used latrine types is a latrine type with a goose neck, which is 88.41%. Furthermore, type of toilet can be divided into types of latrines which are categorized as proper sanitation and improper sanitation. The types

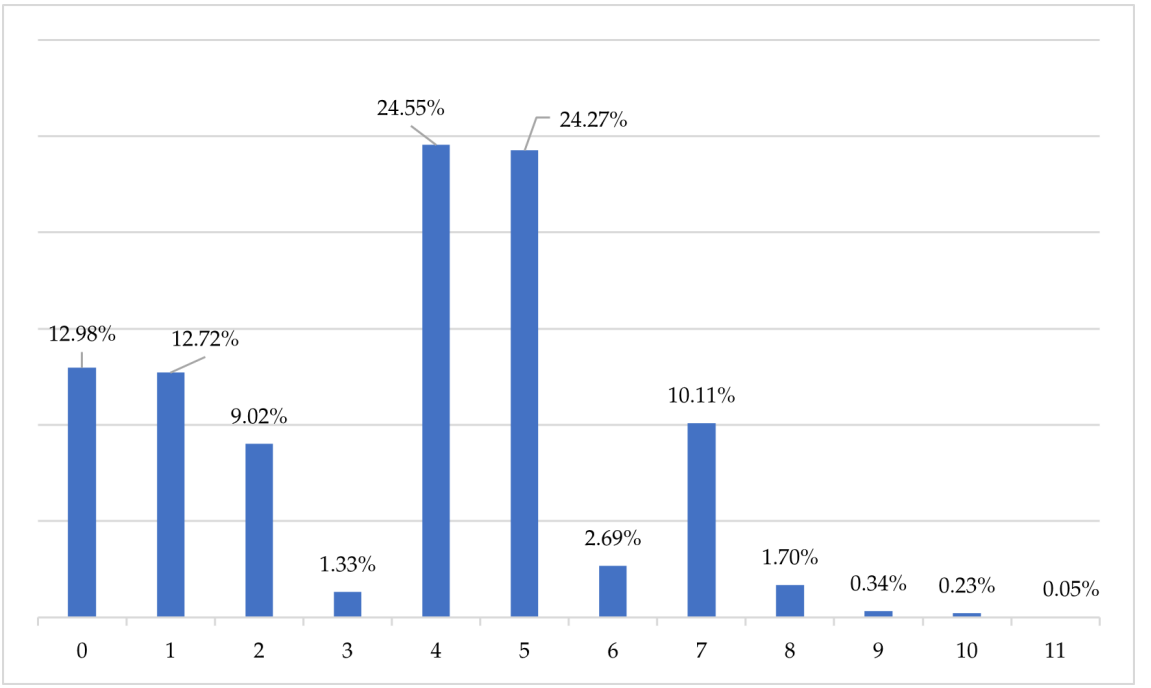

Figure 1: Percentage of main drinking water source of household in East Java with 12 categories.

Information:

0: Branded bottled water	6: Unprotected well
1: Refill Water	7: Protected spring
2: Meter blast	8: Unprotected springs
3: Retail boom	9: Surface Water
4: Drilling well/pump	10: Rain Water
5: Well protected	11: Etc.

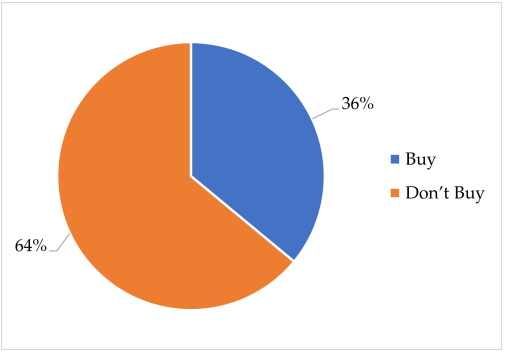

Figure 2: Percentage of main drinking water source of household in East Java with 2 new categories.

of latrines included in the category of proper sanitation are types of goose neck latrines, while for the other three types of latrines (plengsengan with lid, plengsengan without lid, and cemplung

or cubluk) are types of latrines categorized into improper sanitation.

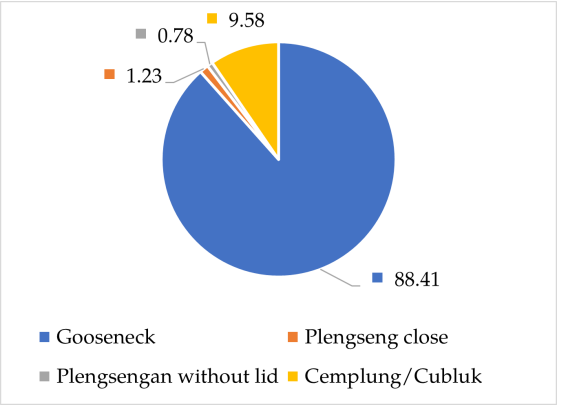

Figure 3: Percentage based on types of latrine used

The percentage of latrines that was classified into proper sanitation (88.41%) and improper sanitation (11.59%) as shown in Figure 4.

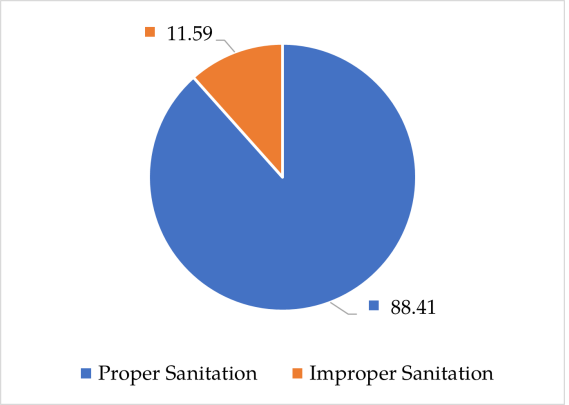

Figure 4: Percentage of latrine types categorized into eligible sanitation and not eligible sanitation

4.2 Grouping Categorical Data

The type of category variables used are the use of PAM water, the drinking water source, the cooking water source, the bathing water source, the distance of the drinking water source with the nearest waste or sewage or feces, as well as the distance from the cooking water source with the nearest waste or feces storage. Grouping for categorical data is by using the ROCK method. The first stage of grouping using the ROCK method is to form a similarity matrix between the objects of observation. The similarity matrix in the ROCK method can explain the distance between observation objects used in research. The similarity value obtained, the value will be between 0 to 1. After obtaining the similarity matrix between the objects of observation, the next step is to

determine the value of θ as the boundary for determining neighbors. After obtaining information on the neighbor relationship between the objects of observation, then the calculation of the number of links and the goodness measure / local heap is carried out. The global heap calculation will then be performed, which is the maximum value of the local heap between columns in i row. After obtaining the global heap value, the objects / clusters that have the maximum global heap value are merged into one group. Then do iteration again starting from the calculation of the similarity matrix until it is found that the object of observation has joined the groups as expected or no link between objects / clusters is found. The theta values in this research is from 0.01 to 0.55, where the optimum grouping results has been chosen from the ratio value and the smallest and there is no outlier in the resulting grouping results. The following are the ratio values and those produced by each value used in this research as shown in the following Table 2.

Table 2: The ratio values of each theta used in ROCK method

Theta	Ratio	Theta	Ratio	Theta	Ratio
0.01	0.00	0.21	**7.04x10^{-2}**	0.41	4.18x10^{-11}
0.02	0.00	0.22	**7.04x10^{-2}**	0.42	4.18x10^{-11}
0.03	0.00	0.23	**7.04x10^{-2}**	0.43	4.18x10^{-11}
0.04	0.00	0.24	**7.04x10^{-2}**	0.44	4.18x10^{-11}
0.05	0.00	0.25	**7.04x10^{-2}**	0.45	4.18x10^{-11}
0.06	0.00	0.26	**7.04x10^{-2}**	0.46	4.18x10^{-11}
0.07	0.00	0.27	0.00	0.47	4.18x10^{-11}
0.08	1.51x10^{-1}	0.28	0.00	0.48	4.18x10^{-11}
0.09	2.09x10^{-1}	0.29	0.00	0.49	4.18x10^{-11}
0.1	1.46x10^{-1}	0.3	0.00	0.5	4.18x10^{-11}
0.11	1.4x10^{-1}	0.31	0.00	0.51	4.18x10^{-11}
0.12	2.23x10^{-1}	0.32	0.00	0.52	4.18x10^{-11}
0.13	1.43x10^{-1}	0.33	0.00	0.53	4.18x10^{-11}
0.14	1.14x10^{-1}	0.34	0.00	0.54	4.18x10^{-11}
0.15	1.27x10^{-1}	0.35	0.00	0.55	4.18x10^{-11}
0.16	1.18x10^{-1}	0.36	0.00		
0.17	**7.04x10^{-2}**	0.37	0.00		
0.18	**7.04x10^{-2}**	0.38	0.00		
0.19	**7.04x10^{-2}**	0.39	0.00		
0.2	**7.04x10^{-2}**	0.4	0.00		

Table 2 shows that the smallest ratio is $4.18x10^{-11}$ on θ from 0.44 to 0.55, but in this θ, there is outliers. When we can found outliers from the result, it same means that the result was not good for representing grouping results, so this theta was not selected. The result of ratio value 0 is not selected, because it produces only one group in the results of the grouping, so it is considered not selected in the research. Taking this into account, the selected smallest ratio value is 7.04x10^{-2} on the θ from 0.17 to 0.26. The grouping result from all this θ (from 0.17 to 0.26) was same, so we can choose one θ value that represents the results of grouping. For this research, the selected theta value is 0.17. The value of this ratio means that the standard deviation in the group is 7.04x10^{-2} times the standard deviation between groups or it can be said that the variance of data in groups gives the value of standard deviation is smaller than the variance between groups.

The results of grouping using the theta 0.17 value are 2 optimum groups, where group 1 is formed by 2 districts or cities, while 36 other districts or cities are classified in group 2. The following are shown by the members of each group in the following Table 3.

Table 3: Continue of the ratio values of each theta used in ROCK method

Group	Districts or Cities
1	Malang City, Madiun City
2	Pacitan District, Ponorogo District, Blitar District, Jember District, Mojokerto District, Madiun District, Magetan District, Bojonegoro Distric, Lamongan District, Bangkalan District, Sumenep District,Trenggalek District, Tulungagung District, Malang District, Lumajang District, Banyuwangi District, Bondowoso District, Situbondo District, Probolinggo District, Pasuruan District, Nganjuk District, Ngawi District, Tuban District, Sampang District, Pamekasan District, Batu City, Kediri District, Jombang District, Gresik District, Kediri City, Blitar City, Probolinggo City, Pasuruan City, Mojokerto City, Sidoarjo District, Surabaya City

Table 4 shows the characteristics of the data in each group based on the independent variable used.

Table 4: Characteristic of categorical data in ROCK

Variable	Category	Group 1	Group 2
The use of PAM water	Don't PAM	50.00	**91.67**
	PAM	50.00	8.33
The main drinking water	Buy	**100.00**	19.44
	Don't Buy	0.00	**80.56**
The main cooking water	Buy	**100.00**	8.33
	Don't Buy	0.00	**91.67**
The main bath/wasting water	Buy	**100.00**	2.78
	Don't Buy	0.00	**97.22**
Distance of drinking water source with waste shelter	$<$ 10 meter	0.00	11.11
	$\geq$10 meter	**100.00**	**88.89**
Distance of cooking water source with waste shelter	$<$ 10 meter	50.00	11.11
	$\geq$10 meter	50.00	**88.89**
Distance of bathing water source with waste shelter	$<$ 10 meter	50.00	13.89
	$\geq$10 meter	50.00	**86.11**

Based on table 4, it can be seen that the districts or cities in group 2 use don't PAM water which is 91.67%, its contrast to group 1, its shows that RTs in the district or city which used don't PAM or PAM are equal which is 50%.

Almost all household in the districts or cities in group 1 get their main drinking water source, the main cooking water source, and the main source of water for bathing or washing by buying, which is 100%. Unlike group 1, most of the household in the districts or cities in group 2 mostly get drinking water sources, sources of cooking water, and the main source of bathing or washing water by not buying.

Distance of drinking water sources with the nearest waste, feces in all districts or cities in groups 1 and 2, has a more dominant distance of more than or equal to 10 meters. In contrast

to these variables, the distance of cooking water source and bath or washing water sources for districts or cities in group 2 has more dominant distance of more than or equal to 10 meters, while for districts or cities in group 1 are equal with 50%. It indicates that the distance of cooking water source and bathing or washing water sources with the nearest waste, feces for districts or cities in group 1 with 50% more than or equal to 10 meters and the other 50% is less than 10 meters.

5 Discussion

Almost all household in the districts or cities in group 1 get their main drinking water source, the main cooking water source, and the main source of water for bathing or washing by buying, which is 100%. Unlike group 1, most of the household in the districts or cities in group 2 mostly get drinking water sources, sources of cooking water, and the main source of bathing or washing water by not buying.

Distance of drinking water sources with the nearest waste, feces in all districts or cities in groups 1 and 2, has a more dominant distance of more than or equal to 10 meters. In contrast to these variables, the distance of cooking water source and bath or washing water sources for districts or cities in group 2 has more dominant distance of more than or equal to 10 meters, while for districts or cities in group 1 are equal with 50%. It indicates that the distance of cooking water source and bathing or washing water sources with the nearest waste, feces for districts or cities in group 1 with 50% more than or equal to 10 meters and the other 50% is less than 10 meters.

6 Conclusions

Based on the discussion of the analysis, obtained 2 optimum groups of results of categorical data with the ROCK method, where group 1 is formed by 2 cities, while 36 other districts or cities are grouped in group 2. It can also be seen that group 1 uses clean water and has better sanitation than group 2. Therefore, it can provide information that member in group 1 are cities that have good access to environmental sanitation. Otherwise, member in group 2 are districts and cities that have poor access to environmental sanitation. This information can also be an information for the East Java Provincial Health Office in making policies and to improve efforts for districts and cities in group 2.

Acknowledgement The authors would like to thank Badan Pusat Statistik (BPS) for supporting this research and providing the data set.

Conflicts of Interest The authors declare no conflict of interest.

References

[1] Dinas Kesehatan Provinsi Jawa Timur (2017). *Profil kesehatan provinsi Jawa Timur*. Kementerian Kesehatan Republik Indonesia, Indonesia.

[2] M. Dutta, A. K. Mahanta & A. K. Pujari (2005). QROCK: A quick version of the ROCK algorithm for clustering of categorical data. *Pattern Recognition Letters*, *26*(15), 2364–2373.

[3] S. Guha, R. Rastogi & K. Shim (2000). ROCK: A robust clustering algorithm for categorical attributes. *Information systems*, *25*(5), 345–366.

[4] L. S. Langit (2016). Hubungan kondisi sanitasi dasar rumah dengan kejadian diare pada balita di wilayah kerja Puskesmas Rembang 2. *Jurnal Kesehatan Masyarakat* (*e-Journal*), *4*(2), 160–165.

[5] E. O. Permatasari & P. Wulandari (2015). Multivariate adaptive regression spline (MARS) approach for classification of sexually transmitted infections (IMS) diseases on commercial sex workers (PSK) in moroseneng localization. *International Journal of Applied Mathematics & Statistics*, *53*(4), 174–180.

[6] L. D. B. Pratiwi, W. Wibowo & I. Zain (2015). Klasifikasi nilai peminat SBMPTN (seleksi bersama masuk perguruan tinggi negeri) ITS dengan pendekatan classification regression trees (CART). *Jurnal Sains dan Seni ITS*, *4*(2), D193–D198.

[7] M. Saleh (2014). Hubungan kondisi sanitasi lingkungan dengan kejadian diare pada anak balita di wilayah kerja Puskesmas Baranti Kabupaten Sidrap tahun 2013. *Jurnal Kesehatan*, *7*(1), 221–233.

[8] A. N. Sidhi, M. Raharjo & N. A. Y. Dewanti (2016). Hubungan kualitas sanitasi lingkungan dan bakteriologis air bersih terhadap kejadian diare pada balita di wilayah kerja Puskesmas Adiwerna Kabupaten Tegal. *Jurnal Kesehatan Masyarakat* (*e-Journal*), *4*(3), 665–676.

[9] A. Tyagi & S. Sharma (2012). Implementation of ROCK clustering algorithm for the optimization of query searching time. *International Journal on Computer Science and Engineering*, *4*(5), 809.

[10] N. N. Veridiana, P. P. F. Sumolang & S. Chadijah (2014). Hubungan pengetahuan, perilaku, dan sanitasi lingkungan dengan angka kecacingan pada anak sekolah dasar di kota Palu. *Media Penelitian dan Pengembangan Kesehatan*, *24*(1), 20695. https://doi.org/10.22435/mpk.v24i1.3487.50-56.