TELKOMNIKA, Vol.15, No.1, March 2017, pp. 504~511
ISSN: 1693-6930, accredited **A** by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v15i1.4750

Potential Usage Estimation of Ground Water using Spatial Association Rule Mining

Suci Sri Utami Sutjipto[1], Imas Sukaesih Sitanggang[2], Baba Barus[3]
[1]Department of Research and Development Technology, Regional Water Company Bogor City, Bogor 16142, Indonesia
[2]Department of Computer Science, Bogor Agricultural University, Bogor 16680, Indonesia
[3]Department of Soil Sciences and Land Resources, Bogor Agricultural University, Bogor 16680, Indonesia
*Corresponding author, email: sc.utami@gmail.com[1], imas.sitanggang@ipb.ac.id[2], bababarus@yahoo.com[3]

Abstract

The utilization of ground water in the long term will lead to a number of negative impacts on groundwater resources and the environment, such as the decrease of groundwater level, seawater intrusion, land subsidence as well as scarcity of ground water. Furthermore, the use of ground water has directly affected the consumption pattern of Regional Water Company Bogor City (PDAM) customers. This study aims to determine the patterns and characteristics of PDAM customers in the utilization of ground water by using spatial association rule mining, so it can help PDAM to approximate the increase of customers that utilize ABT and the losses incurred. This research shows that as many as 53.362 (41.27%) PDAM customers that have the potential to use groundwater. The said customers are featured by several characteristics, such as being active customers, with monthly water bill of less than Rp. 53.358 and are not close to river.

Keywords*: ground water, spatial association rule, apriori algorithm*

1. Introduction

Underground water is the water that fills in the void in the geological layer or saturated zone (also known as ground water). The water in the saturated zone is important for engineering works, geological study as well as developing water supply [1]. The exploitation of ground water to meet household and commercial drink water demand has recently grown significantly, which goes along with the increasing population growth and development activities. However, excessive use of ground water may lead to ground water crisis, particularly the underground water. Unless this situation is immediately addressed, it is very likely that more severe impacts -water scarcity, among other- might occur. According to the data from the Environmental Management Agency (BPLH) [2], the ground water extraction from legal well (well with permit) in Bogor City from 2012 until 2014 has increased by more than 100%; that is, from 606.354 m^3 to 1.339.572 m^3. The data of the ground water extraction through legal well (Figure 1).

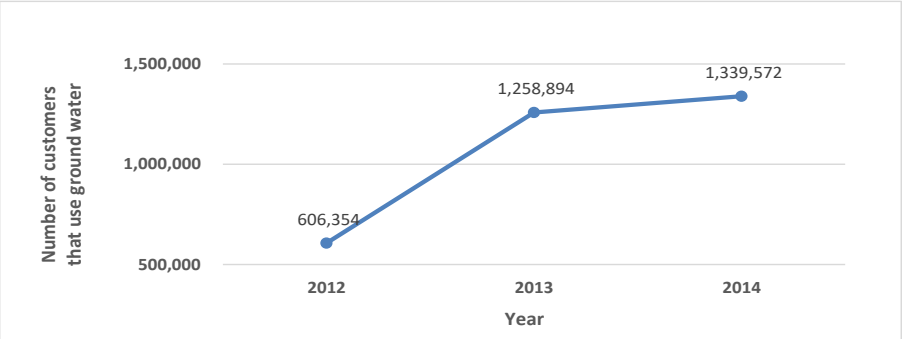

Figure 1. Number of Customers that use Ground Water through Legal Well

Received September 11, 2016; Revised December 7, 2016; Accepted December 24, 2016

The ground water utilization must be accompanied with appropriate control and conservation efforts, since if there were a decline in ground water quality, it would take a fairly long time to recover. One of the control processes which can be conducted is by using the Local Government Water Company (PDAM) as the water provider to the community. However, up to date [3], some of PDAM customers are identified as ground water users, and until the end of December 2014, the number of PDAM Tirta Pakuan of Bogor City customers were 129.312 active customers (including 2.872 customers in Bogor District, which are served by PDAM Tirta Pakuan of Bogor city). This shows that 78.41% of the total population of Bogor City, which constituted 956.760 people in 2014, used PDAM water, which indicates that the issue of ground water is not only related to environmental one, but might also lead to another impact, i.e. the decline of PDAM customer (Figure 2), which eventually will affect the company's revenue.

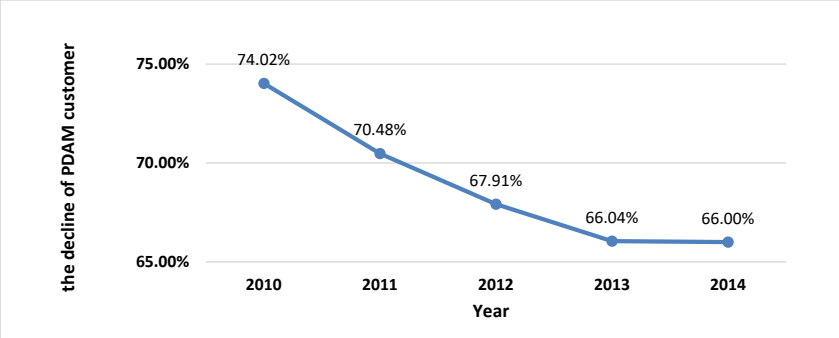

Figure 2. Water usage of Customers of PDAM of Bogor City

To figure out the trend of community's usage of ground water, an analysis on the data pattern of ground water usage at every keypoint is required. Keypoint is a coordinate point indicating the location of ground water users. The keypoint data is obtained from the Environmental Management Agency of Bogor City and PDAM of Bogor City. Attributes found at the keypoint such as customer ID and customer location will be utilized to find out the using pattern and distribution of groundwater ownership. Mapping the pattern will expectedly lead to the identification of which attributes affecting the community's trend in using the ground water, particularly the ones who are also PDAM customers.

Extracting usage pattern and associative relations for large data can be conducted by using data mining approach. Data mining is a process of extracting the needed knowledge from large data [4]. In the process, data mining will extract valuable information by analyzing patterns or certain relations from the large data. One of the most commonly used techniques to find association pattern of a pool of data is the spatial association rule mining, which is the extension of association rule mining [5-7].

The first thing conducted to find out the association rules is to find frequent itemsets, which is a group of items that often occur at the same time. After all the frequent itemsets are found, associative rules that have met the set requirements will be figured out. The algorithm that is often used to find associative rules is the apriori algorithm.

The search of pattern using apriori algorithm has already been widely developed, including in the study by [8] entitled "The association rule mining for ground water and wastelands using apriori algorithm: case study of Jodhpur District". The study analyzed wastelands which contained a lot of ground water in Jodhpur area by using the association Rule Mining method by implementing the apriori algorithm. The study outcome shows that wasteland that contains a lot of ground water in Jodhpur District can be found in Bilara area.

In this study, an analysis test on ground water source ownership against PDAM water use will be conducted by using the spatial association rule mining to figure out the usage pattern of PDAM customers utilizing ground water. The data source to be used is the PDAM customer data, comprising both users and non-users of ground water as well as the data of ground water user community obtained from BPLH. Spatial-wise, association is an inter-connection between one spatial object with another spatial object, which in this regard refers to the connection

between ground water user customer with the customer's location (settlement) as well as the supporting attributes.

The goal to be achieved in this study is to determine the characteristics of customers utilizing ground water in order to analyze the ground water use potential among the customers of PDAM Tirta Pakuan of Bogor City by using the Spatial Association Rule Mining approach. This study is expected to bring about several benefits, including the identification of the common pattern of customers using ground water, as well as estimation of how many more customers that might utilize ground water in order to be able to address the decrease in use pattern.

2. Research Method

The study conducted consists of three major steps (Figure 3), namely: the spatial data pre-process, application of apriori algorithm to obtain the association pattern of ground water source ownership, and the analysis of ground water usage potential among the customers of PDAM Tirta Pakuan of Bogor City.

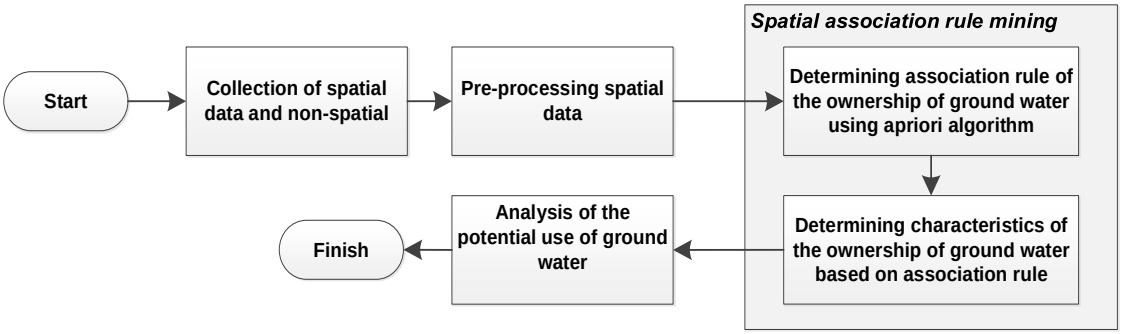

Figure 3. Research Steps

3. Results and Analysis

3.1. Spatial and Non-Spatial Data Collection

The data used in this study consists of spatial data (Figure 4) and non-spatial data which is obtained from the query of Customer Information System (CIS) of PDAM of Bogor City. The attributes found in the non-spatial data include customer ID, sub-district, village, rate class, customer status, average water use in m^3, average cost of use, and remarks.

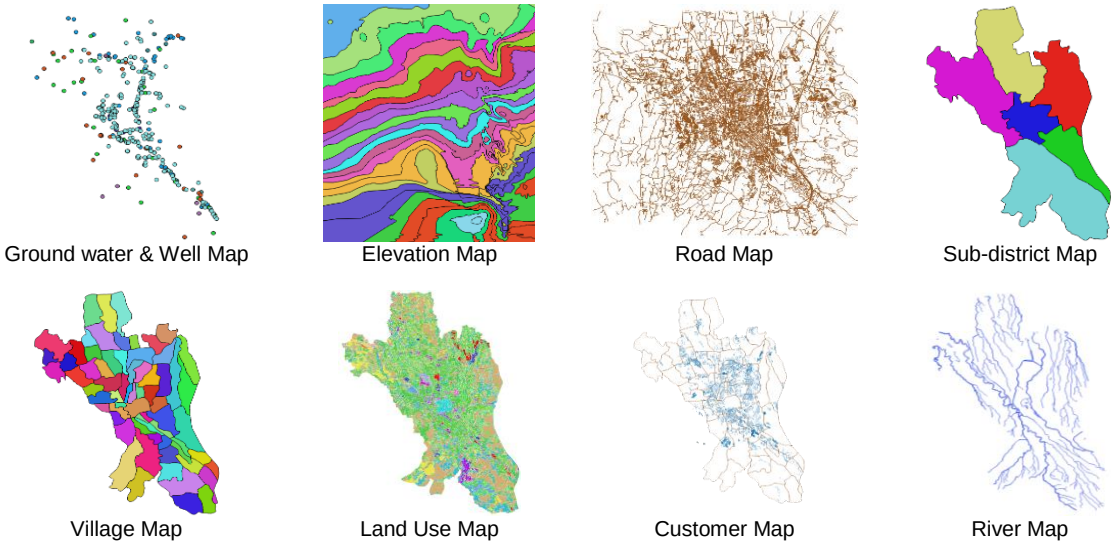

Figure 4. Spatial Data used in this Study

3.2. Spatial Data Pre-Processing

After the data collection is completed, the next step is data pre-processing. There are two different sets of data, namely spatial and non-spatial data, which will constitute two data sets, firstly the spatial data set, and secondly the non-spatial data set. To merge both data sets into the spatial data base, the spatial data are processed by using queries and spatial operations. No spatial data obtained from the CIS query will be selected and integrated through Kettle Pentaho, to be subsequently put into the spatial data base.

The result generated from the spatial data pre-processing steps is a spatial database which consists of lake which consists of 4 polygon features, wells with 91 point features, groundwater with 316 point features, road with 17,086 line features, river with 415 line features, landuse with 16,912 polygon features, customer with 80,003 polygon features, village with 68 polygon features, sub-district with 6 polygon features, and customers with 85,221 data. On the other hand, the result of spatial operation process generated 566 records with 10 variables, including distance from the road (near_road) with buffer value of 5m, distance from river (near_river) with buffer value of 30m, ground water spot, elevation, land use, water use, water fee, village, sub-district, customer status and customer tariff class.

3.3. Association Rule Mining using Apriori Algorithm

After the spatial data pre-processing step is completed, the next step is algorithm application which aims to obtain the association pattern of ground water source ownership, in order to figure out the caracteristics of ground water user customer. Apriori algoritm is a basic algoritm introduced by [9] to define the frequent itemsets for boolean association rules. The rule that declares association between several attributes is commonly referred to affinity analysis or market basket analysis. Apriory algorithm uses knowledge on the frequent itemset that has been previously known to process the subsequent information. Apriori algorithm is used to determine the possibly occurring candidates by noting the minimum support.The two main processes conducted in the apriori algorithm [4] are:

1 *Join* (merging process). In this process, each item is combined with another item until no more combination may be generated. $\boldsymbol{C_k}$ (itemset candidate with k size) is generated by combining $\boldsymbol{L_{k-1}}$ (*itemset* that frequently occurs in k size).
2 *Prune* (cutting-off). In this process, the combined item is then pruned by using minimum support that has been set by user. In this regard, the minimum support used is 80%. Thus, the item set that does not frequently occur as a part (k-1) will be pruned. Hence the apriori algorithm application [4]:

Apriori ($\boldsymbol{T}, \boldsymbol{\varepsilon}$)
$\boldsymbol{L_1}$ ← { large 1-itemsets that appear in more than $\boldsymbol{\varepsilon}$ transaction}
$\boldsymbol{k} \leftarrow \boldsymbol{2}$
while $\boldsymbol{L_{k-1}} \neq \emptyset$
$\boldsymbol{C_k}$ ← Generate ($\boldsymbol{L_{k-1}}$)
For transaction $\boldsymbol{t \in T}$
$\boldsymbol{C_t} \leftarrow$ Subset ($\boldsymbol{C_k}, \boldsymbol{t}$)
For candidate $\boldsymbol{c \in C_t}$
$\boldsymbol{count}[\boldsymbol{c}] \leftarrow \boldsymbol{count}[\boldsymbol{c}] + \boldsymbol{1}$
$\boldsymbol{L_k} \leftarrow \{ \boldsymbol{c \in C_k} | \boldsymbol{count}[\boldsymbol{c}] \geq \boldsymbol{\varepsilon} \}$
$\boldsymbol{k} \leftarrow \boldsymbol{k} + \boldsymbol{1}$
Return $\cup_{\boldsymbol{k}} \boldsymbol{L_k}$

The result of apriori algorithm application is the characteristics of PDAM customers which also use ground water. The characteristics will be used for evaluation process to identify the pattern of ground water using customers.

3.4. Evaluation of Ground Water User Customer Characteristics

Association rule is considered as an interesting pattern if it meets the minimum benchmark value for each metric. The example of association rule is as follows [4]: (item1, item2) → (item3) (support=40%, confidence=50%). From the association rule, 50% of the transaction in the database consists of item1 and item2 as well as item3. As many as 40% of the entire transaction consists the three items. The support value is the measure of how

frequent an item or itemset appears in the entire transaction. Support from rule A→ B can be calculated by using the following formula [4]:

$$\text{support} = (A \Rightarrow B) = P(A \cup B) = \frac{\text{number of transaction using item A and B}}{\text{number of entire transaction}} \quad (1)$$

To calculate confidence value [4] which is the measure that shows the relation between both items according to certain conditions, which in this regard the measure from association A→ B, the following formula can be applied [2]:

$$\text{confidence} = (A \Rightarrow B) = P(A|B) = \frac{\text{number of transaction containing item A and B}}{\text{number of transaction containing item A}} \quad (2)$$

Other than support and confidence calculation, in applying apriori algorithm, there is also a lift calculation [10]:

$$\text{lift} = (A \Rightarrow B) = \frac{P(A \cup B)}{P(A \times B)} \quad (3)$$

Spatial Association Rule Mining is the extension of Association Rule Mining by using spatial data. The association rule in the spatial data is stated in the following format [11]:

$x1 \wedge x2 \wedge \ldots \wedge xm \rightarrow y1 \wedge y2 \wedge \ldots \wedge yn$ (sup%, con%)

The above format shows the associative relations between predicate xi (i=1,…,m) and yj (j=1,…,n), where there is at least one spatial predicate. The example of spatial association rule mining is as follows [11]:

is_a(X,sumur) /\close_to(X,0-20)/\depth(X,0-250)/\inside(X,basin14)
→ arsenic_level(X, classlabel:dangerous) (20%, 80%)

This rule identifies that 80% of the wells are close to factory within 20 km distance, and the river is located in basin14 with depth of less than 250 feet, and contain arsenic with dangerous concentration level; 20% meet the three predicates in the above.

In this study, the apriori algorithm application aims to identify the spatial association rules implemented by using the R statistic application (https://www.rstudio.com/). The goal of this study is to find out the characteristics of customers using ground water, so as to be able to predict how many more customers that might potentially also use ground water. Therefore, for further analysis, only the rule that has ground water or has_ground water = “yes” that will be used. Application of the apriori algorithm produces 597 rules on the minimum support of 10% and minimum confidence of 80%. Scatter plots for 597 (Figure 5) association rules which contain has_ground water=yes. Every dot in the scatter plot represents rules, which in this regard refers to the characteristics of customers using ground water. Support and lift are used for x-axis and y-axis where the color in points represents the level of confidence.

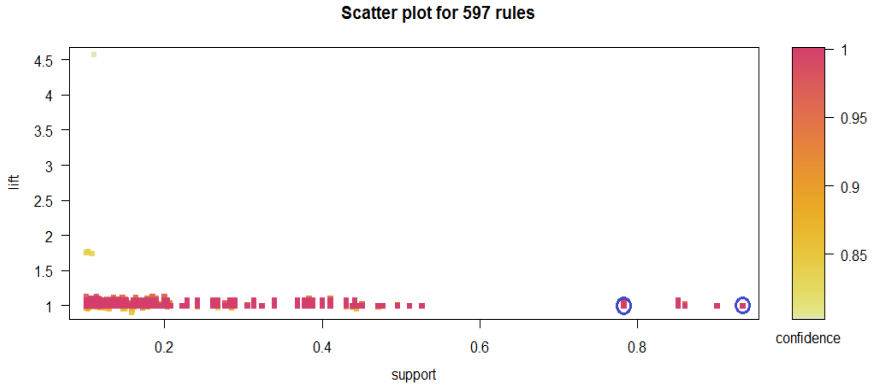

Figure 5. Scatter Plot 597 of Association Rules that have Predicate of has_ground Water = Yes

On Figure 5, the bottom right dot (blue circle on the rightmost) has the following rule:

	lhs	rhs	support	confidence	lift
Rule 1:	{inv_kategori=0-53358}	=> {has_abt=yes}	0.9336438	1	1

Rule 1 means 93.36% of customers using groundwater has characteristics monthly water bill of less than Rp. 53.358, with confidence level of 100%. The example of another rule is:

	lhs	rhs	support	confidence	lift
Rule 2:	{near_river=no, inv_kategori=0-53358, status_kategori=status3}	=> {has_abt=yes}	0.7826931	1	1

Rule 2 indicates that 78.26% with confidence level of 100% of customers that use ground water have the characteristics of active customers, with monthly water bill of less than Rp. 53.358 and are not close to river. Other rules generated from the apriori algorithm application is shown on Table 1.

Table 1. Rules Generated From Apriori Algorithm Application

Rules	*support*	*conf*	*lift*
{inv_kategori=0-53358} => {has_abt=yes}	93.36%	100%	1.00
{near_river=no} => {has_abt=yes}	90.10%	100%	1.00
{near_river=no,inv_kategori=0-53358} => {has_abt=yes}	86.03%	100%	1.00
…	…	…	…
{near_river=no,status_kategori=status3} => {has_abt=yes}	78.27%	100%	1.00
{near_river=no,inv_kategori=0-53358,status_kategori=status3} => {has_abt=yes}	78.27%	100%	1.00
…	…	…	…
{near_road=yes,near_river=no,pel_elevasi=240-250,status_kategori=status3} => {has_abt=yes}	10.01%	100%	1.00
{near_road=yes,near_river=no,pel_elevasi=240-250,inv_kategori=0-53358,status_kategori=status3} => {has_abt=yes}	10.01%	100%	1.00

3.5. Analysis of Potential Usage of Ground Water

The rules that are made as the basis for evaluating the pattern of customers having ground water are those with the confidence level of ≥60% and ≥80%, as it truly influences the confidence of prediction level which determines the possibility of customer who use ground water, which serves as the basis to evaluate customers who are identified using ground water, to predict whether such customers have the possibility of utilizing ground water. The rules with minimum support of 60% are as follows:

1. {near_river=no,inv_kategori=0-53358} => {has_abt=yes}
2. {status_kategori=status3} => {has_abt=yes}
3. {inv_kategori=0-53358,status_kategori=status3} => {has_abt=yes}
4. {near_river=no,status_kategori=status3} => {has_abt=yes}
5. {near_river=no,inv_kategori=0-53358,status_kategori=status3} => {has_abt=yes}

The rules with minimum support of 80% are as follows:

1. {inv_kategori=0-53358} => {has_abt=yes}
2. {near_river=no} => {has_abt=yes}
3. {near_river=no,inv_kategori=0-53358} => {has_abt=yes}
4. {status_kategori=status3} => {has_abt=yes}
5. {inv_kategori=0-53358,status_kategori=status3} => {has_abt=yes}

The selected rules are subsequently used to determine the potential of groundwater use by PDAM customers, which will be afterwards put into the dataset indicating customers who don't have ground water. Based on the rules of minimum support of 60% and a minimum support of 80%, this study obtains 53.362 (41.27%) PDAM customers that have the potential to use groundwater. The said customers are featured by several characteristics, such as being active customers, with monthly water bill of less than Rp. 53.358 and are not close to river.

PDAM customers that have the potential to use groundwater based on the rules of minimum support of 60% and minimum support of 80% are mostly distributed in several villages, including Bantarjati Village (4.186 customers), Baranangsiang Village (3.019 customers), Empang Village (2.044 customers), Curug Mekar Village (1.869 customer), Katulampa Village (1.628 customers), Cibogor Village (1.421 customers), Bondongan Village (1.212 customers), Menteng Village (1.150 customers), Pasir Jaya Village (1.067 customers) and Gudang Village (1.024 customers). The distribution of PDAM customers which might potentially use ground water (Figure 6).

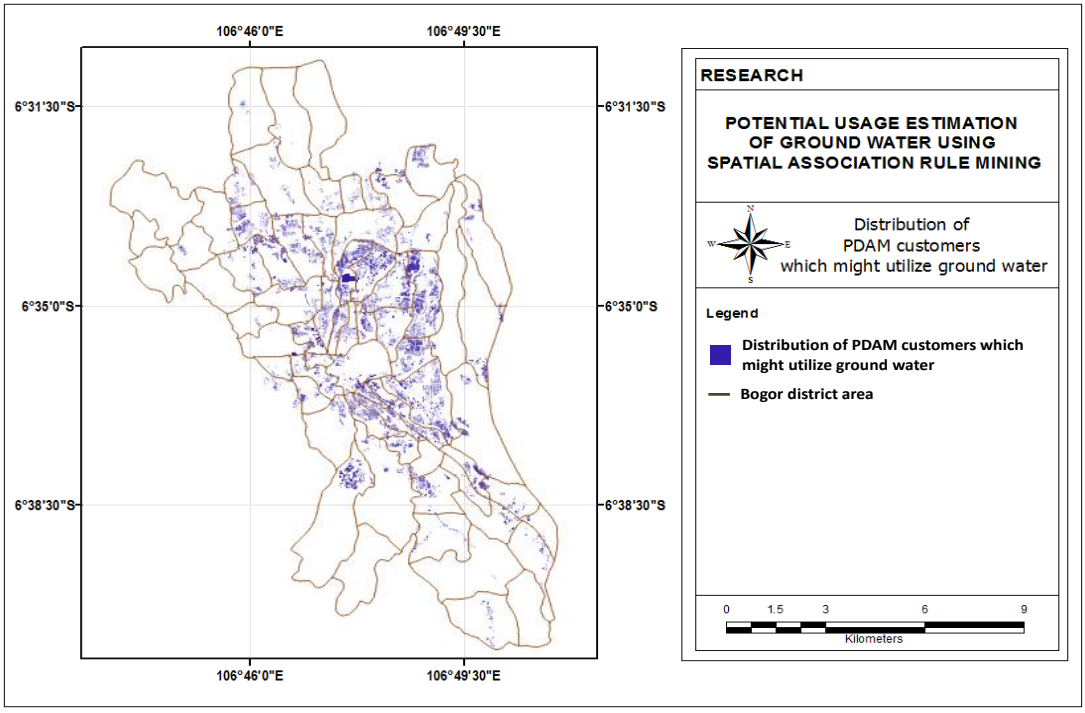

Figure 6. Distribution of PDAM Customers which Might Utilize Ground Water

This study indicates that further study is needed by including not only the customers of PDAM Tirta Pakuan of Bogor City, but also broader community. The results of this study can be also used in further evaluation for PDAM Tirta Pakuan of Bogor City to estimate loss of revenue which the company would suffer due to the declining use of customers as they move from PDAM to ground water. In addition, this study can be used in decision making by PDAM Tirta Pakuan of Bogor City, particularly as the basis for reclassifying the tariff group, evaluating tariff increase, as well as being a company's review material in socializing and promoting the use of PDAM water instead of ground water in effective mannern. In addition, this study can be further developed for social purpose, i.e. to assist in developing government policy to determine the ground water retribution tax, which would make the community choose PDAM water. Finally, this study can assist related parties in developing government policy concerning with environmental issue that needs to be thoroughly studied for its long-term impact of ground water usage.

4. Conclusion

Application of the apriori algorithm produces 597 rules on the minimum support of 10% and minimum confidence of 80%. Based on the rules of minimum support of 60% and a minimum support of 80%, this study obtains 53.362 (41.27%) PDAM customers that have the potential to use groundwater. The said customers are featured by several characteristics, such as being active customers, with monthly water bill of less than Rp. 53.358 and are not close to river.

PDAM customers that have the potential to use groundwater based on the rules of minimum support of 60% and minimum support of 80% are mostly distributed in several villages, including Bantarjati Village (4.186 customers), Baranangsiang Village (3.019 customers), Empang Village (2.044 customers), Curug Mekar Village (1.869 customer), Katulampa Village (1.628 customers), Cibogor Village (1.421 customers), Bondongan Village (1.212 customers), Menteng Village (1.150 customers), Pasir Jaya Village (1.067 customers) and Gudang Village (1.024 customers).

References

[1] Todd DK, Mays LW. 2005. Groundwater Hydrology. Hoboken (US): John Wiley & Sons, Inc.

[2] Bogor BPLHK. 2011. Penyusunan Rencana Induk Terpadu Pengelolaan Air Tanah Kota Bogor. Bogor (ID): Badan Pengelolaan Lingkungan Hidup Kota Bogor.

[3] Badan Litbang PDAMTPKB. 2014. Laporan Tahunan Badan Penelitian dan Pengembangan PDAM Tirta Pakuan Kota Bogor. Bogor (ID): PDAM Tirta Pakuan Kota Bogor.

[4] Han J, Kamber M, and Pei J. 2012. Data Mining: Concepts and Techniques. San Fransisco (US): Elsevier, Inc.

[5] Man M, Bakar WAWA, Abdullah Z, Jalil MA, Herawan T. 2016. Mining Association Rules: A Case Study on Benchmark Dense Data. *Indonesian Journal of Electrical Engineering and Computer Science*. 3(No. 3, September 2016): 546-553

[6] Wang Z. 2013. An Efficient Association Rules Algorithm Based on Compressed Matrix. *TELKOMNIKA*. 11(No. 10, October 2013): 5711-5717

[7] Agrawal R, Imieliński T, Swami A. 1993. Mining Association Rules Between Sets of Items in Large Databases. Di dalam: ACM SIGMOD International Conference; 1993 Jun 01; New York (US): ACM. Hlm 207-216.

[8] Faridi M, Verma S, Mukherjee S. 2015. Association Rule Mining for Ground water and Wastelands Using Apriori Algorithm: Case Study of Jodhpur District. *International Journal of Advanced Research in Computer Science and Software Engineering.* 5(6): 751-758

[9] Agrawal R, Srikant R. 1994. Fast Algorithms for Mining Association Rules. San Jose (CA): IBM Almaden Research Center.

[10] Sergey B, Motwani R, Ulman JD, Tsur S. *Dynamic Itemset Counting and Implication Rules for Market Basket Data.* Proceedings ACM SIGMOD International Conference on Management of Data. 1997: 255-264

[11] Koperski K, Han J. Discovery of Spatial Association Rules in Geographic Information Databases. Di dalam: 4th International Symposium, SSD'95; 1995 Agustus 06–09; Portland (US): Springer Berlin Heidelberg. 1995: 47-66.