

PAPER • OPEN ACCESS

Monitoring Application for Clean Water Access and Clustering using K-Means Algorithm

To cite this article: N R Radliya *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **662** 022096

View the [article online](#) for updates and enhancements.

Monitoring Application for Clean Water Access and Clustering using K-Means Algorithm

N R Radliya¹, M R Fachrizal², A R Rabbi³

^{1,2}Departemen Sistem Informasi, Universitas Komputer Indonesia

³Departemen Pendidikan Lingkungan Hidup, Universitas Siliwangi

Email: nizar@email.unikom.ac.id

Abstract. The purpose of this research is to develop a monitoring application of clean water access for the community and determining areas that need to improve access to clean water. Data collection was done by examining documents which were related to clean water management and conducting interviews with the Department of Housing and Settlement of West Java Province. The results of data collection are used to develop a clean water monitoring application system based on Android and Web. Data from the application will be used for clustering with the *K-Means* algorithm. For example, in the city of Bandung, 12 urban village areas were recommended for increasing access to clean water. The area has a percentage of ownership of clean water access approaching the value of 29.96% of the ownership of clean water sources for food and drink, 28.96% for ownership of other sources of clean water and 14.48% for ownership of reserves of clean water. The optimal number of clusters of clean water access in the city of Bandung is three clusters because it has the lowest *Davies-Bouldin Index* (DBI) value of 0.439.

1. Introduction

Clean water is good quality water that can be used by the community for consumption or daily activities, including sanitation [1]. The needs of clean water for the district/city community continue to increase along with population growth in each area. The fulfillment of clean water continues to be improved by the Indonesian government, precisely at the provincial level, by gradually developing access to clean water in each area [2]. The development of a clean water supply system in principle aims to create clean water quality management and services that are of high quality, continuity to the public at an affordable price, so as to balance the interests between the community and clean water service providers [3].

This research was conducted at the Department of Housing and Settlement of West Java Province. The government is still constrained in getting information regarding the extent to which the community needs to access clean water [4]. This information can only be obtained when conducting data collection directly in the field. Based on the idea that the data obtained must reflect as much as possible the activities that occur in the field, then the solution to the solution can be done by sending data from the field of monitoring through an effective and efficient application. Information on the current condition of clean water access will facilitate the government in determining areas that need to increase access to clean water to maintain public health in the region [5].

The purpose of this research is to develop a monitoring application of clean water access for the community and determining areas that need to improve access to clean water. Data collection is done



by examining documents, which are related to clean water management and conducting interviews with the Department of Housing and Settlement of West Java Province. In this research, an application for monitoring access to clean water was carried out which aims to determine the extent of the area of the community whose clean water needs have been met. The implementation of clean water access monitoring system is carried out from hierarchically from data sources in the regular monitoring of community facilitators or authorized officials sent to the district/city level and routed to the province on a regular basis every time period. The application was developed using android and website based technology. The monitoring application developed will produce baseline data regarding ownership of clean water access for each community or area. The next baseline data will be used for the clustering process which aims to group between areas that have been met with access to clean water and those who have not had access to clean water. This will be useful for the government in developing access to clean water. The clustering process uses the *K-Means* algorithm. *K-Means* algorithm is an algorithm used for data/object grouping techniques based on the distance of each data [6]. The software used for *K-Means* algorithm implementation is *RapidMiner* [7]. In this software, clustering modeling will be made for the recommendation system for developing access to clean water in several areas.

2. Methods

The research procedure was based on the Research and Development (RandD) method combined with the system approach method and the system development life cycle method. A distinctive feature of the RandD method was the research method used for research that produces products [8]. In this research would produce products in the form of applications and clustering models. The system development life cycle method used in this study was the Rational Unified Process (RUP). The RUP method is presented in Figure 1.

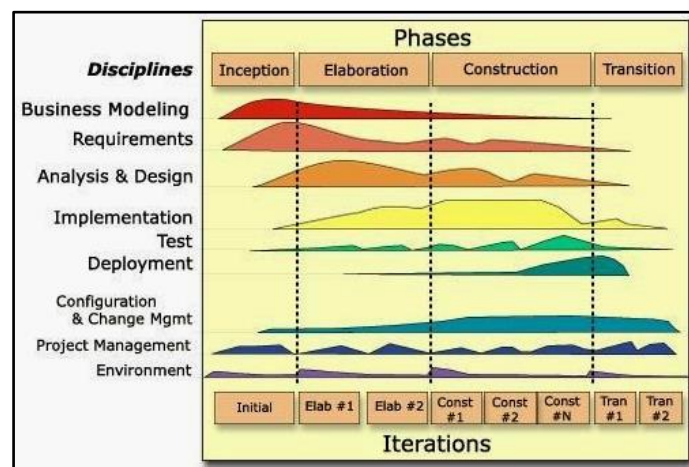


Figure 1. Rational Unified Process (RUP) [9]

The RUP method is an iterative software development process that has two dimensions of the work phase [9]. The first dimension is drawn horizontally consisting of inception, elaboration, construction, and transition. Each phase in the first dimension is done iteratively. The second dimension is described vertically consisting of business modeling, requirements, analysis and design, implementation, test, deployment, configuration and change management, project management, environment. Each activity in the second dimension is carried out in each phase of the first dimension in accordance with the predetermined portion. The procedure of research in this study is presented in Figure 2.

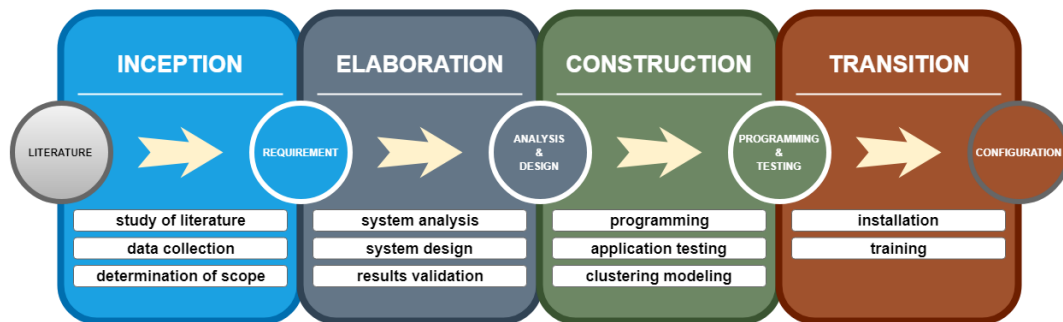


Figure 2. Research procedure

2.1. Phase I: Inception

At this phase, a preliminary study was conducted to determine the scope of the research. At this phase data collection is related to clean water access. Data collection is done by means of literature studies and data collection in the field through interviews and document analysis [10]. The results of this phase are the scope of the functionality of the monitoring application and the objectives of clustering modelling.

2.2. Phase II: Elaboration

The elaboration phase is the preparation stage before entering the construction phase. At this stage, an analysis of the data collection system for clean water access is currently being carried out by the West Java provincial government. The results of the analysis will be used as a reference for designing clean water access monitoring applications. Analysis and design are documented using modelling tools namely Unified Model Language (UML). UML is a set of structures and techniques for object-oriented program design modelling [11].

2.3. Phase III: Construction

At this phase, programming is implemented based on a validated design. The application developed will be tested and revised to produce applications that have functionality in accordance with the initial provisions. Data will be used for making clustering models. Making clustering models using *RapidMiner* software tools.

2.4. Phase IV: Transition

The transition phase is the stage for the dissemination and application of the application system to the research subject. At this stage, training in the use of monitoring applications and the use of clustering models will be conducted.

3. Results and Discussion

3.1. System Analysis

Clean water sources refer to the Central Bureau of Statistics (BPS) in Indonesia with the WHO (World Health Organization). Some sources of clean water are presented in Table 1.

Table 1. Clean water access

Clean Water Source	BPS	WHO
Well protected	feasible	feasible
Open well	not feasible	not feasible
Boreholes	feasible	feasible
Water terminal	feasible	feasible
Springs protected	feasible	feasible
Open spring	not feasible	not feasible
PDAM	feasible	feasible
Rainwater reservoir	feasible	feasible
Gallon water/packageing water	not feasible	not feasible

Water around/dirigent not feasible not feasible

The proportion of households with sustainable access to clean water is the ratio between households with access to clean water sources and all households expressed as a percentage [12]. Households who use clean water are those who use clean water to meet the life needs of their members. The percentage of households using clean water in West Java during the period 2012-2017 is presented in Table 2.

Table 2. Households of clean water users in 2012-2017 [12]

<u>Year</u>	<u>Percentage</u>
2012	60.31
2013	63.50
2014	65.01
2015	67.20
2016	68.81
2017	71.57

From Table 2 it can be seen that the number of households using clean water each year continues to increase. This is the impact of the Community Based Water Supply and Sanitation Program (Pamsimas). Pamsimas is a national flagship program (government) to increase population access to clean water and sanitation facilities with a community-based approach [4].

The implementation of the clean water access assessment was carried out at the urban village level by involving the elements within it (cadres) as actors of data recording at the community level. For the urban village that already have drinking water and sanitation management organizations such as BPSPAMS, the data collection actors can be carried out by BPSPAMS independently while still coordinating with the urban village. The organizational structure implementing the data on access to clean water is shown in Figure 3.

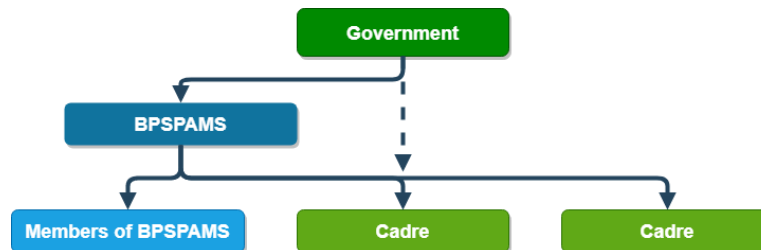


Figure 3. The implementing organizational structure for the data collection on clean water [13]

3.2. System Design

The use case diagram serves to show the interaction between the system and the actor/user [11]. Modelling the application monitoring design uses a use case diagram in Figure 4.

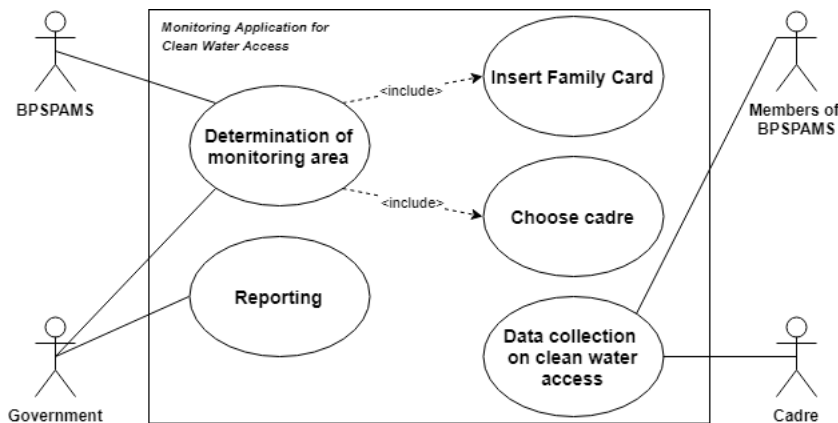


Figure 4. Use case diagram of application for monitoring access to clean water

The use case description of the clean water access monitoring application is presented in Table 3.

Table 3. Use case description

Use Case	Description	Actor
Determination of monitoring area	Determine the area where data on access to clean water will be conducted. Determination of area in each urban village is based on the smallest entity at that level.	BPSPAMS, Government
Insert Family Card	Input family card data for each area that has been previously determined.	BPSPAMS, Government
Choose cadre	Determine officers who will collect data on clean water access for each area that has been determined.	BPSPAMS, Government
Data collection on clean water access	Collect data in the field regarding ownership of clean water access for each community.	Members of BPSPAMS, Cadre
Reporting	Check the status of fulfilling clean water access in each area.	Government

3.3 Development of Monitoring Application

The use of the application for monitoring clean water access begins with regional data collection carried out by the provincial government. Regional data collection is done through a website application that is connected to the android application. The appearance of the web-based monitoring application is presented in Figure 5.

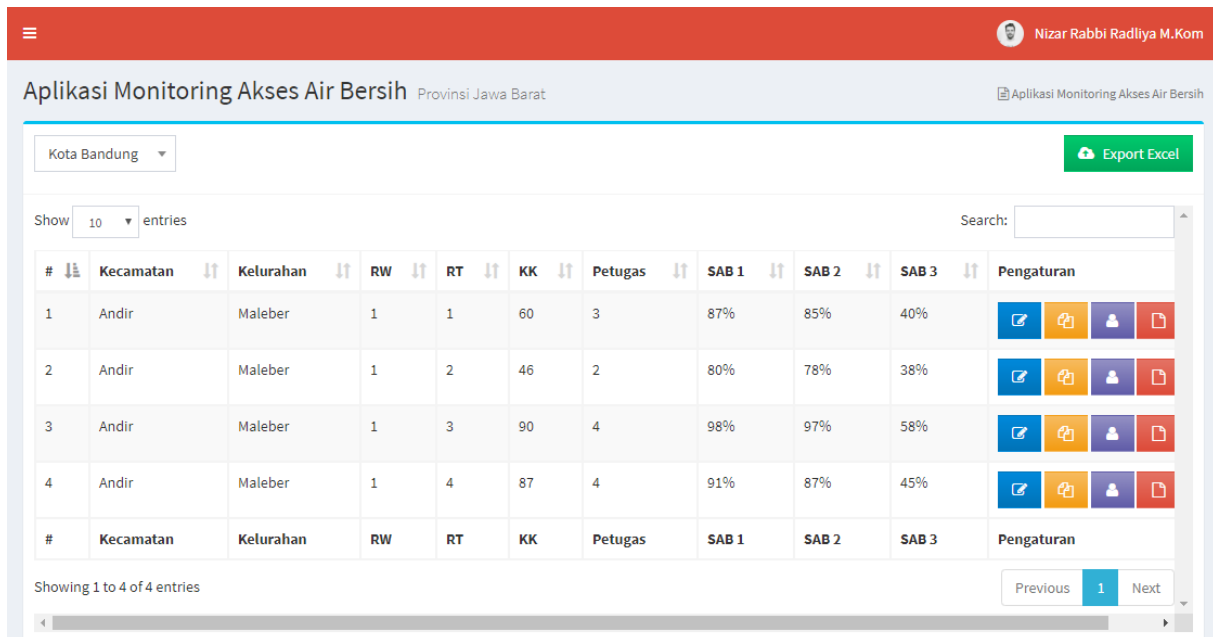


Figure 5. The interface of monitoring application web-based for clean water access

The urban village level government (BPSPAMS or urban village head) registers as a monitoring coordinator in each area using an android application. The coordinator must determine which data on access to clean water will be conducted. Each area must also have a number of families (based on family cards) and officers who will collect data on access to clean water. Data collection officers (members of BPSPAMS or cadres) who have been given the task can collect data on clean water access based on the area that has been determined. The results of the data collection can be monitored in real-time by the government. The appearance of the android-based monitoring application is presented in Figure 6.



Figure 6. The interface of monitoring application android-based for clean water access

3.4. Clustering Modelling

The main purpose of clustering techniques is to group data into a number of groups (clusters) based on the characteristics of the data [14]. The stages of the *K-Means* algorithm are presented in Figure 7.

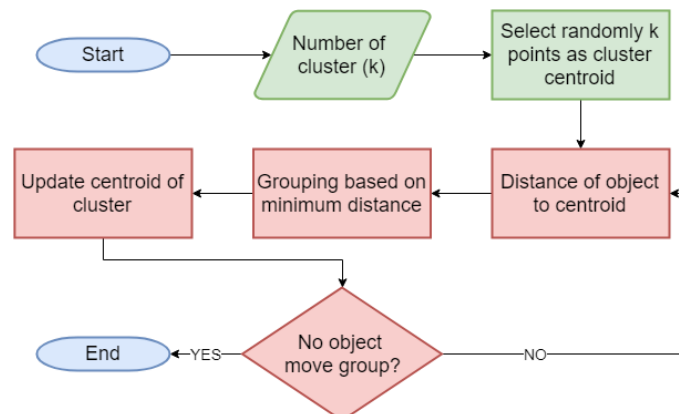


Figure 7. *K-Means* algorithm

The *K-Means* algorithm begins by determining the number of clusters (*k*) and determining the centroid value randomly for each cluster. Next, we calculate the distance of each object with centroid using the euclidian formula shown in formula (1).

$$D(x,y) = \sqrt{\sum_{j=1}^n |x - y|^2} \tag{1}$$

Each object will be allocated to the nearest centroid. After that, determine the new centroid value by calculating the average value of objects in each group. Calculation of the distance of each object to the centroid will continue until no object acts on the cluster [15].

The monitoring data will be used for the clustering modelling process. Clustering modelling uses the *K-Means* algorithm with *RapidMiner* software tools. *K-Means* clustering modelling is shown in Figure 8.

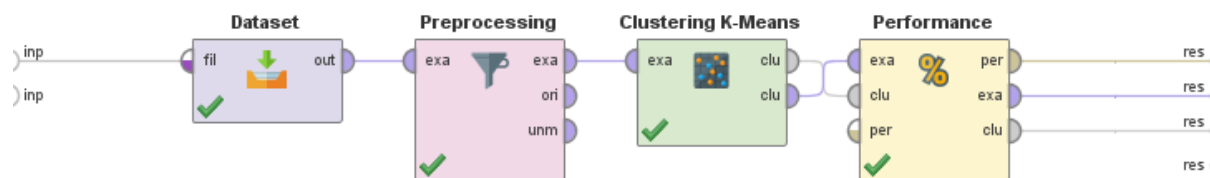


Figure 8. The process of clustering modeling for clean water access

The dataset used for the clustering process is the percentage of data that fulfills access to clean water at the urban village level. The features of the dataset include sources of clean water for cooking and drinking water sources. For example, a dataset is used to monitor clean water access in Bandung, which has 120 sub-districts (urban village). The pre-processing is used to filter data with missing value. The pre-processing produces 113 rows of data that will enter the k-means clustering stage. The k-means clustering process produces three clusters as in Table 4.

Table 4. The results of clustering clean water access of Bandung city

Cluster	Centroid 1	Centroid 2	Centroid 3	Number of Items	Recommendation
Cluster 1	88.34	87.39	43.67	64	Low
Cluster 2	61.39	60.39	30.19	37	Medium
Cluster 3	29.96	28.96	14.48	12	High

All urban villages included in cluster three are the priority areas recommended for obtaining clean water access development programs. As for the area of the sub-district which is included in the cluster of three, the area is Ciroyom, Maleber, Babakan, Cibangkong, Kebon Waru, Samoja, Jamika, Suka Asih, Mekar Wangi, Cisurupan, Rancabolang, Sindang Jaya.

Block performance functions to determine the quality of the number of clusters based on the *Davies-Bouldin Index* (DBI). The DBI value of each number of clusters is presented in Table 5.

Table 5. Value of *Davies-Bouldin Index* (DBI)

2 Clusters	3 Clusters	4 Clusters	5 Clusters	6 Clusters
0.539	0.439	0.479	0.449	0.532

Case of clean water access of Bandung city, the most optimal number of clusters is three. DBI is a cluster validation method from clustering results. The DBI measurement approach is to maximize the inter-cluster distance and minimize intra-cluster distance. The smaller the DBI value shows the most optimal cluster scheme [16].

4. Conclusion

The monitoring application can serve to monitor the condition of community ownership of clean water access. The government can more easily collect data on clean water access currently. Data generated from this application can be used for the process of grouping ownership of clean water access in each area. For example, in Bandung city, 12 urban village areas were recommended for the development of clean water access. The area has a percentage value of ownership of clean water access approaching the value of 29.96% of ownership clean water sources for food and drink, 28.96% for ownership other sources of clean water, and 14.48% for ownership reserves of clean water. The optimal number of clusters of clean water access of Bandung city is three clusters because it has the lowest *Davies-Bouldin Index* (DBI) value of 0.439.

Acknowledgement

Thanks to UNIKOM that have been supported and presented in INCITEST 2019. Members of the Department of Information are helped for completing research. Especially thanks to Chancellor of UNIKOM Dr. Ir. H. Eddy Soeryanto Soegoto, MT has remembered me about the milestone report and for all support provided.

References

- [1] Lilje, J., and Mosler, H. J. (2017). Socio-psychological determinants for safe drinking water consumption behaviors: a multi-country review. *Journal of Water, Sanitation and Hygiene for Development*, **7**(1), pp.13-24.
- [2] Patunru, A. A. (2015). Access to Safe Drinking Water and Sanitation in Indonesia. *Asia and the Pacific Policy Studies*, **2**(2), pp.234-244.
- [3] Willetts, J., Gero, A., Susanto, A. A., Sanjaya, R., Trieu, T. D., Murta, J., and Carrard, N. (2017). Sanitation value chains in low density settings in Indonesia and Vietnam: impetus for a rethink to achieve pro-poor outcomes. *Journal of Water, Sanitation and Hygiene for Development*, **7**(3), pp. 445-454.
- [4] Pratama, A. B., and Isnani, A. T. (2018). Evaluasi Berjalan Program Penyediaan Air Minum dan Sanitasi Berbasis Masyarakat. *Jurnal Ilmu Administrasi: Media Pengembangan Ilmu dan Praktek Administrasi*, **15**(2), pp.148-162.
- [5] A. Wahyudi and A. A. Patunru (2019). "Sanitation, Drinking Water Access and Diarrhea in Indonesia," *Russian Journal of Agricultural and Socio-Economic Sciences*, **85**(1), pp.142-152.
- [6] Ahmar, A. S., Napitupulu, D., Rahim, R., Hidayat, R., Sonatha, Y., and Azmi, M. (2018, June). Using K-Means Clustering to Cluster Provinces in Indonesia. In *Journal of Physics: Conference Series* **1028**(1), pp. 012006. IOP Publishing.
- [7] Hofmann, M., and Klinkenberg, R. (Eds.). (2013). *RapidMiner: Data mining use cases and*

- business analytics applications*. CRC Press, **12**(2), pp. 122-232.
- [8] Sugiyono (2010) *Metode Penelitian Kuantitatif, Kualitatif, RandD*, Bandung: Alfabeta, **4**(1), pp.456-501.
- [9] R. S. Pressman and B. Maxim (2014) *Software Engineering: A Practitioner's Approach*, 8th Ed, New York: McGraw-Hill, **5**(3), pp.365-423.
- [10] M. Nazir (2009). *Metode Penelitian*, Bogor: Ghalia Indonesia, **3**(2), pp.85-125.
- [11] K. E. Kendall and J. E. Kendall (2013). *Systems Analysis and Design*, 9th ed, New Jersey: Pearson Prentice Hall, **4**(1), pp.243-342.
- [12] Badan Pusat Statistik (BPS) (2017). "Survei Sosial Ekonomi Nasional (SUSENAS)," Jakarta, **7**(2), pp.234-353.
- [13] C. Chaerunnissa (2015). "Partisipasi Masyarakat Dalam Program Penyediaan Air Minum dan Sanitasi Berbasis Masyarakat (PAMSIMAS) di Kabupaten Brebes (Studi Kasus Desa Legok dan Desa Tambakserang Kecamatan Bantarkawung)," *Politika: Jurnal Ilmu Politik*, **5**(2), pp. 99-113.
- [14] Zaki, M. J., Meira Jr, W., and Meira, W. (2014). *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press, **6**(2), pp.253-273.
- [15] Han, J., Kamber, M., and Pei, J. (2011). *Data mining concepts and techniques third edition*. *The Morgan Kaufmann Series in Data Management Systems*, **5**(4) pp.83-124.
- [16] Maulik, U., and Bandyopadhyay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on pattern analysis and machine intelligence*, **24**(12), pp.1650-1654.