

The Application of Clustering Technique to Water Quality of Surabaya River

Sri Rahmawati F. ^{a,*}, Mohammad Isa Irawan ^a, Nieke Karnaningroem ^b

^a Institut Teknologi Sepuluh Nopember, Faculty of Mathematics and Natural Science Department of Mathematics, Jalan Raya ITS, Surabaya, Indonesia

^b Institut Teknologi Sepuluh Nopember, Department of Environmental Engineering, Jalan Raya ITS, Surabaya, Indonesia

* Corresponding author e-mail address: rahmawati13@mhs.matematika.its.ac.id

Abstract

Surabaya river is the source of the raw water used people in Surabaya to satisfy their daily needs. Surabaya river flow starts from Mlirip (Mojokerto) as upstream past the Sidoarjo, Gresik, and until the downstream that is Jagir Wonokromo (Surabaya). The background research studies because the water surface in Surabaya decreased perceived water quality is constantly increasing as a result most of the liquid industrial waste discharged from human activities into the channel that empties in Surabaya either directly or indirectly. Water pollutant components Surabaya river is known biological oxygen demand known as BOD, chemical oxygen demand known as COD, total suspend solid known as TSS, and dissolved oxygen known as DO. Even for biological oxygen demand components at some point status monitor heavy polluted with concentrations exceeding waterquality class sub-cleanness. K-means is one of a method for clustering objects based on their characteristics. Object of study in this research is a point source the disposal of industrial wastewater empties in Surabaya river in 2013. The first step of this research study is the normalization of water quality data biological oxygen demand, chemical oxygen demand, total suspend solid and dissolved oxygen in Surabaya river. The results of this step is the concentration of data which are in the range of 0 and 1. Concentrations were normalized value applied to K-means resulting in a model that descibes the shape of the distribution of pollutants proximity region group. From the result obtained K-means suitability of water quality of each region along the Surabaya river is formed of similarity and attribute similarity. Davies Bouldin Index is an internal evaluation scheme, for cluster validation of K-means. From the results of the cluster validation, Davies Bouldin Index values obtained for 53.742. DBI value obtained is minimal DBI to the number of 672 iterations. This paper not only theoretically situation, the water quality of Surabaya river by clustering technique, but also get a conclusion of pollutant with five cluster becomes the best value Davies Bouldin Index.

Keywords: K-means clustering, Water quality, Surabaya river

1. Introduction

Cities play an important role in regional, national and even international development (Zhang and Yang, 2007). Surabaya river is one of surface water that used by Surabaya people to satisfy the daily needs (Nurul et al., 2012). Progressive industrial development has also increasing use of rivers forwaste disposal activities. The pollution from these and the others sources, such as use agricultural pesticides, has led to the increasing need for rigorous assessment of river quality. Water pollution caused serious harm on human's lives, production activities and health.

Determination of the water quality is traditionally based on classification by considering physical, chemical, biological factors, and heavy metals (Su et al., 2012) according to the purpose of water usage (Ay and Kisi, 2012).

Because modelling of the water parameters is an important and complex issue, different Artificial Intelligence (AI) techniques are performed in modelling for

various water resources areas (Irawan et al., 2013). In this context, there are several parameters to assess water quality according to the national and international criteria standards. These parameters include biological oxygen demand, chemical oxygen demand, dissolved oxygen, and total suspend solids. According to the data report from Perum Jasa Tirta (PJT) of East Java province, for the last five years from 2009 to 2013 (Rahmawati et al., 2014a).

For instance, the AI techniques have been studied by several research in modelling rainfall, evaporation, discharge, BOD, COD and DO concentration, depth integrated DO and sediment over the last three decades. Based on this knowledge, these techniques have also been succesfully used in estimation and forcasting (Zeleňáková et al., 2012) of water resources. For instance, Artificial Neural Network (ANNs) were identified, tested and validation for the computation DO (Ay and Kisi, 2012) and BOD (Singh et al., 2009). When using ANN method to comprehensively assess air quality using RBF network by classifying pollutants in assessment grades from air

component, RBF network was an effective assessment to judge the combined effects of pollutants on air (Jie et al., 2014).

Clustering techniques is also one of the other data analysing techniques. This techniques allow retrieval of the useful information by grouping or categorizing multi-dimensional data in cluster. Moreover, it is very important for data mining system and in desicion making process (Zirnea et al., 2013). Therefore, clustering techniques has been successfully used in various areas such as artificial intelligence, pattern recognition, engineering, economy, geology, electronics, statistic, psychology and marketing. For instance, Rahmawati et al (2014b) analysed a 3-year for 2011 until 2013 sediment of BOD, COD and DO used the kohonen self-organizing maps for estimation of concentration of Surabaya river in Surabaya.

The main purpose of clustering techniques is to partitionate a set of entities into different groups, called clusters (Zirnea et al., 2013). These group may be consistent in term of similarity of its members. As the name suggests, the representative-based clustering techniques use some form of representative for each cluster and every group has a member that represents it. K-means use the strategy of representative based clustering to water quality of Surabaya river.

2. Natural Geographical Situation

Surabaya river is downstream tributary of Brantas River, the longest river in East Java which has an area of the watershed 630.7 km², consisting of 289.7 km² watershed Kali Marmoyo, watershed area of 99.4 km² Kali Watudakon, and some of the tributary watershed covering an area 227.3 km² (Surabaya, 2014).

3. Research Methods

3.1 Cluster Analysis

Cluster analysis is mainly used to study classification of various object or phenomena. The basic idea of this approach is to set each samples a class of their own, and then define the distance between the samples (or a similar factor) and the distance between classes. Select from the smaller pair, merge into a new class, and calculate the new class and the distance between the other classes, and the merging the closest two. This will reduce the number of the classes, one at a time, until all of the samples are distributed into one same class. According to the definition of between-class distance method, which is divided into the shortest distance, the distance method, the middle distance, center of gravity method. Group average method, variable group average method (Zirnea et al., 2013).

K-means clustering technique is one of the clustering techniques and has been used in many areas. This techniques is based on distance matrix (Hartigan and Wong, 1979). Euclidean distance is used as s distance criterion. Algorithm starts with k initial seeds of clustering. All n data ate then compared with each sees by means of the

Euclidean distance and are assigned to the closest cluster seed. The method is then repeated again and again until the error becomes minimum value. The accuracy of the k-means procedure is very dependent on the choice of the initial seed (Milligan and Cooper, 1985).

Silhouette plot method (Rousseeuw, 1987) refers to a method of interpretation and validation of clusters of data. This technique provides a concise graphical representation of how well each object within its cluster. Eq. (1) is the objective function; j_i is the Euclidean distance between i data vector (x_k) and j clsuetr centre (c_i); c is the number of cluster ($c \in N$ and $2 \leq c \leq n$), and u_{ij} is i membership degree. These parameters are calculated with Eqs. (4) and (5) according to the conditions (3).

$$J = \sum_{i=1}^c J_i = \sum_{i=1}^c (\sum_{k, x_k \in G_i} \|x_k - c_i\|^2) \quad (1)$$

$$U = \begin{bmatrix} u_{11} & \cdot & u_{1n} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ u_{c1} & \cdot & u_{cn} \end{bmatrix}_{c \times n} \quad (2)$$

After cluster centres are calculated with Eq. (3), the membership degress (1 or 0) are determined according to Eq. (2) by minimizing Eq. (1), which is objective function

$$u_{ij} = \begin{cases} 1 & \text{if } \|x_j - c_i\|^2 \leq \|x_k - c_i\|^2 \\ 0 & \text{if } \|x_j - c_i\|^2 > \|x_k - c_i\|^2 \end{cases}; \text{ for } k \neq i \quad (3)$$

$$|G_i| = \sum_{j=1}^n u_{ij} \quad (4)$$

$$c_i = \frac{1}{|G_i|} \sum_{k, x_k \in G_i} x_k \quad (5)$$

3.2 K-Means Algorithm

The K-means is a greedy, computationally efficient technique, being the most popular representative-based clustering algorithm (Hartigan and Wong, 1979).

The algorithm is composed of the following steps (Li and Wu, 2012) :

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all object hae been assigned, recalculate the positions of the K centroids.
4. Repeat Step 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

K-means clustering is a partitioning method that treats observations in your data as objects having locations and distances from each other. It partitions the objects into K mutually exclusive clusters, such that objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. Each cluster is

characterized by its centroid, or center point (Li and Wu, 2012).

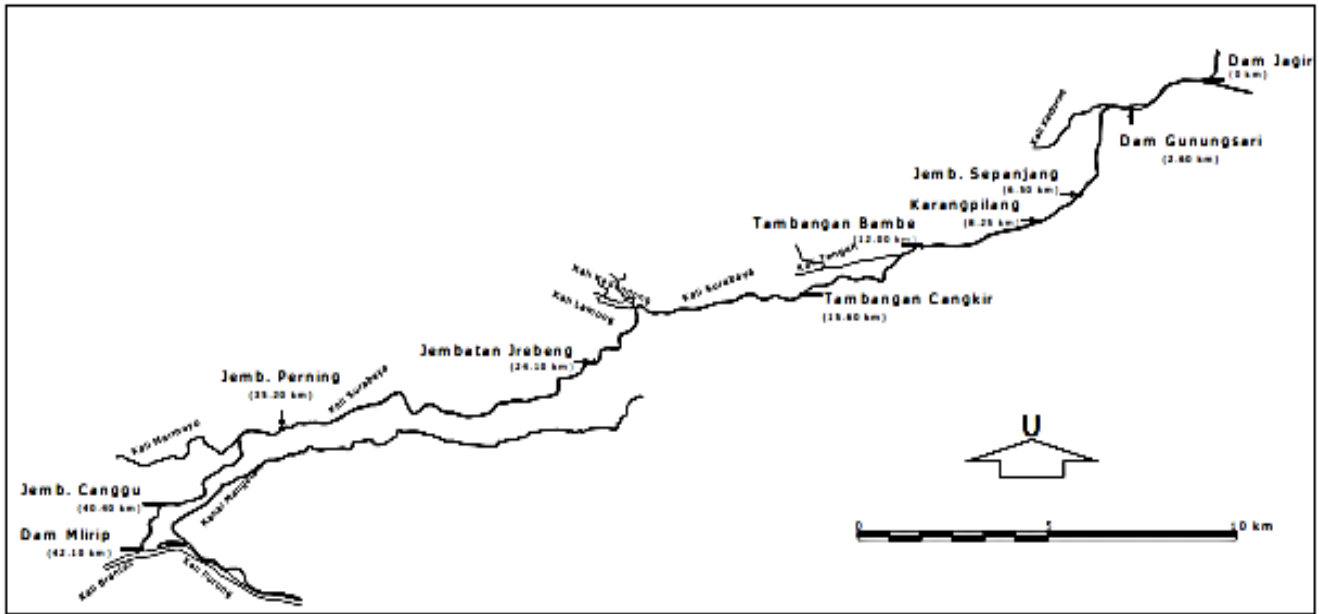


Fig. 1. Location and sampling sites in Surabaya river

3.3 Davies Bouldin Index

The Davies Bouldin Index (DBI) introduced by David L. Davies and Donald W. Bouldin in 1979. The Davies-Bouldin Index is a metric for evaluating clustering algorithm. This is an internal evaluation scheme, where the validation of how well the clustering has been done is made using quantities and features inherent to the data set.

Let $R_{i,j}$ be a measure of how good the clustering scheme is. This measure, by definition has to account for $M_{i,j}$ the separation between i and j cluster, which ideally has to be as large as possible, and S_i the within cluster scatter for cluster i , which has to be as low as possible. Hence (Davies et al, 1979) the Davies-Bouldin Index is defined as the ratio of S_i and $M_{i,j}$ such that these properties are conserved :

1. $R_{i,j} \geq 0$.
2. $R_{i,j} = R_{j,i}$
3. If $S_j \geq S_k$ and $M_{i,j} = R_{i,k}$ then $R_{i,j} > R_{i,k}$
4. And if $S_j = S_k$ and $M_{i,j} \leq M_{i,k}$ then $R_{i,j} >$

$R_{i,k}$

$$R_{i,j} = \frac{S_i + S_j}{M_{i,j}} \tag{6}$$

This is the symmetry condition. Due to such a formulation. The lower the value, the better the separation of the clusters and the tightness inside teh clusters.

$$D_i \equiv \max_{j:i \neq j} R_{i,j} \tag{7}$$

If N is the number of clusters:

$$DBI = \frac{1}{N} \sum_{i=1}^N D_i \tag{8}$$

Tabel 1

Standart of surface water quality classification (Zhao et al., 2012).

Comprehensive pollutions index	Water quality level
≤ 0.20	I Cleanness
0.21 - 0.40	II Sub-cleanness
0.41 - 1.00	III Slight pollution
1.01 - 2.0	IV Moderate pollution
≥ 2.01	V Severe pollution

4. Result and Discussion

Using clustering algorithm based on K-Means cluster, water quality of 33 sampling site in Surabaya river from January to December 2013 were comprehensive assessed. Davies Bouldin Index, final cluster, distance between cluster were processed by Matlab tools.

Normalization is a pretreatment method to sample data befor K-Means assessment. All data are transformed to the interval of [0,1] or [-1,1] by normalizing. By doing this, magnitudes differences of dimension werw eliminated so that statistical distribution of sample could be unified and the convergence of training network was usually be accelerated. Maximum and minimum method for normalizing was shown as Eq. (9). We applied all for clustering algorithms on the real data sets describe below.

$$x'_k = \frac{x_k - x_{min}}{x_{max} - x_{min}} \tag{9}$$

Table 3 shows the Davies Buoldin Index value resulting from the application for two cluster until nine cluster. The lower the value, the better the separation of the clusters.

The outcomes of algorithm K-Means are describe by Table 4 to Table 7. The outcome of standart K-Means initial cluster focal point are too random. Table 4 to Table 7

gives useful information about the grouping process performed by clustering methods.

Table 2

Davies Bouldin Index.

Output cluster	Iteration history	DBI value
2	209	209.285
3	139	353.452
4	374	164.302
5	672	53.742
6	263	113.37
7	773	87.917
8	196	202.856
9	536	79.743

Table 3

Final cluster for biological oxygen demand.

Sampling site	Membership cluster			
	Jan - March 2013	Apr - June 2013	July - Sept 2013	Oct - Des 2013
Site 1	3	4	5	2
Site 2	3	4	4	2
Site 3	3	4	5	2
Site 4	3	4	5	2
Site 5	3	4	5	2
Site 6	3	4	5	5
Site 7	3	4	5	2
Site 8	3	2	3	5
Site 9	3	4	5	2
Site 10	3	4	5	2
Site 11	3	4	5	2
Site 12	2	2	3	5
Site 13	3	2	5	5
Site 14	2	2	5	4
Site 15	3	3	5	5
Site 16	2	3	4	5
Site 17	1	2	4	2
Site 18	2	4	4	2
Site 19	2	2	4	4
Site 20	3	3	4	2
Site 21	5	2	4	5
Site 22	2	3	4	3
Site 23	3	4	5	2
Site 24	2	4	5	2
Site 25	1	1	1	2
Site 26	2	2	4	5
Site 27	4	3	5	1
Site 28	4	5	2	4
Site 29	5	3	5	5
Site 30	3	2	5	2
Site 31	3	3	5	5
Site 32	3	4	5	2
Site 33	3	4	5	2

Table 4

Final cluster for chemical oxygen demand.

Sampling site	Membership cluster			
	Jan - March 2013	Apr - June 2013	July - Sept 2013	Oct - Des 2013
Site 1	3	4	5	1
Site 2	3	4	2	1
Site 3	3	4	5	1
Site 4	3	4	5	1
Site 5	3	4	5	1
Site 6	3	4	5	2
Site 7	3	4	5	1
Site 8	3	4	5	1
Site 9	3	4	5	1
Site 10	3	4	5	1
Site 11	3	4	5	1
Site 12	3	4	5	1

Table 5

Final cluster for chemical oxygen demand (Cont.)

Sampling site	Membership cluster			
	Jan - March 2013	Apr - June 2013	July - Sept 2013	Oct - Des 2013
Site 13	4	4	2	1
Site 14	3	4	5	1
Site 15	3	4	5	1
Site 16	3	4	5	1
Site 17	4	4	5	1
Site 18	3	1	5	1
Site 19	3	4	5	2
Site 20	3	4	5	4
Site 21	5	4	5	1
Site 22	3	5	5	3
Site 23	3	4	5	1
Site 24	1	4	5	1
Site 25	2	3	1	1
Site 26	3	4	2	1
Site 27	3	5	4	5
Site 28	3	2	3	2
Site 29	5	4	5	1
Site 30	3	4	5	1
Site 31	3	4	5	1
Site 32	3	4	5	1
Site 33	3	4	5	1

Table 6

Final cluster for total suspend solids.

Sampling site	Membership cluster			
	Jan - March 2013	Apr - June 2013	July - Sept 2013	Oct - Des 2013
Site 1	5	4	5	1
Site 2	5	4	4	1
Site 3	3	5	5	1
Site 4	3	4	5	3
Site 5	5	5	5	1
Site 6	3	4	3	3
Site 7	3	5	4	1
Site 8	5	4	5	1
Site 9	5	4	4	1
Site 10	5	5	5	1
Site 11	5	5	5	1
Site 12	5	4	5	5
Site 13	2	4	2	5
Site 14	5	4	5	3
Site 15	5	4	5	5
Site 16	5	4	5	3
Site 17	1	4	5	1
Site 18	5	3	4	4
Site 19	5	4	5	1
Site 20	3	4	5	2
Site 21	3	4	5	5
Site 22	5	2	5	4
Site 23	5	2	5	1
Site 24	5	4	5	1
Site 25	4	1	3	5
Site 26	5	4	2	1
Site 27	3	2	2	4
Site 28	5	2	1	5
Site 29	3	4	5	5
Site 30	5	4	5	5
Site 31	5	2	5	5
Site 32	5	5	5	1
Site 33	3	5	5	1

Table 8 shows the final partition of five cluster, which are the average distance from centroid across all set data biological oxygen demand. Table 9 shows the final partition of five cluster, which are the average distance from centroid across all set data chemical oxygen demand. Table 10 shows the final partition of five cluster, which are

the average distance from centroid across all set data total suspend solids. Table 11 shows the final partition of five cluster, which are the average distance from centroid across all set data dissolved oxygen.

Table 7

Final cluster for total suspend solids.

Sampling site	Membership cluster			
	Jan - March 2013	Apr - June 2013	July - Sept 2013	Oct - Des 2013
Site 1	5	4	1	4
Site 2	5	4	4	4
Site 3	5	5	1	4
Site 4	5	4	1	4
Site 5	5	4	1	4
Site 6	5	4	1	5
Site 7	5	4	1	4
Site 8	5	5	2	5
Site 9	5	5	1	4
Site 10	5	4	1	4
Site 11	5	4	4	4
Site 12	1	5	2	5
Site 13	3	5	2	5
Site 14	1	5	4	5
Site 15	5	1	4	5
Site 16	1	1	4	5
Site 17	3	5	4	4
Site 18	1	1	4	4
Site 19	1	1	4	3
Site 20	1	1	4	2
Site 21	4	5	4	3
Site 22	1	1	4	1
Site 23	5	4	1	4
Site 24	2	4	1	4
Site 25	3	2	3	4
Site 26	1	5	4	5
Site 27	2	1	3	3
Site 28	2	3	5	5
Site 29	2	5	1	5
Site 30	1	5	4	5
Site 31	5	1	1	5
Site 32	5	4	1	4
Site 33	1	4	1	4

Table 8

Final partition of five cluster all set data biological oxygen demand.

	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster 1	0.037	0.135	0.135
Cluster 2	0.506	0.403	0.504
Cluster 3	1.101	0.571	0.841
Cluster 4	0.050	0.076	0.102
Cluster 5	0.022	0.030	0.075

Table 9

Final partition of five cluster all set data chemical oxygen demand.

	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster 1	0.101	0.053	0.174
Cluster 2	0.554	0.409	0.574
Cluster 3	1.012	0.534	0.800
Cluster 4	0.000	0.014	0.014
Cluster 5	0.008	0.065	0.065

Table 10

Final partition of five cluster all set data total suspend solids.

	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster 1	0.274	0.146	0.261
Cluster 2	0.781	0.387	0.698
Cluster 3	0.015	0.067	0.085
Cluster 4	0.961	0.458	0.733
Cluster 5	0.039	0.051	0.117

Table 11

Final partition of five cluster all set data dissolved oxygen.

	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster 1	0.006	0.056	0.056
Cluster 2	0.818	0.508	0.633
Cluster 3	0.140	0.083	0.083
Cluster 4	0.179	0.112	0.256
Cluster 5	0.056	0.055	0.148

Table 12, Table 13, Table 14, and Table 15 represent the centroid cluster of biological oxygen demand, chemical oxygen demand, total suspend solids, and dissolved oxygen.

Table 12

Centroid cluster BOD of the k-means clustering.

Month (2013)	C-1	C-2	C-3	C-4	C-5	Grand centroid
Jan-March	0.880	0.326	0.043	0.066	0.008	0.107
Apr-June	0.031	0.139	0.370	0.031	0.011	0.061
July-Sept	0.111	0.831	0.051	0.076	0.013	0.110
Oct-Des	0.023	0.136	0.580	0.099	0.018	0.100

Table 13

Centroid cluster COD of the k-means clustering.

Month (2013)	C-1	C-2	C-3	C-4	C-5	Grand centroid
Jan-March	0.028	0.231	0.023	0.005	0.971	0.102
Apr-June	0.023	0.508	0.418	0.013	0.016	0.102
July-Sept	0.016	0.756	0.006	0.007	0.063	0.085
Oct-Des	0.026	0.234	0.592	0.192	0.016	0.106

Table 14

Centroid cluster TSS of the k-means clustering.

Month (2013)	C-1	C-2	C-3	C-4	C-5	Grand centroid
Jan-March	0.038	0.342	0.150	0.047	0.015	0.078
Apr-June	0.022	0.050	0.012	0.290	0.013	0.054
July-Sept	0.146	0.712	0.052	0.252	0.038	0.180
Oct-Des	0.182	0.206	0.043	0.633	0.031	0.172

Table 15

Centroid cluster DO of the k-means clustering.

Month (2013)	C-1	C-2	C-3	C-4	C-5	Grand centroid
Jan-March	0.984	0.447	0.086	0.097	0.025	0.151
Apr-June	0.116	0.666	0.312	0.158	0.033	0.158
July-Sept	0.147	0.885	0.091	0.090	0.233	0.137
Oct-Des	0.036	0.127	0.926	0.114	0.015	0.118

Table 16 shows the distance between cluster centroid biological oxygen demand of clustering generated by K-Means.

Table 16

Distance between cluster centroid of biological oxygen demand.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster 1	0.000	0.922	1.063	0.818	0.877
Cluster 2	0.922	0.000	0.969	0.815	0.895
Cluster 3	1.063	0.969	0.000	0.590	0.670
Cluster 4	0.818	0.815	0.590	0.000	0.115
Cluster 5	0.877	0.895	0.670	0.115	0.000

Table 17 shows the distance between cluster centroid chemical oxygen demand of clustering generated by K-Means.

Table 17

Distance between cluster centroid of chemical oxygen demand.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster 1	0.000	0.931	0.690	0.168	0.944
Cluster 2	0.931	0.000	0.861	0.926	1.147
Cluster 3	0.690	0.861	0.000	0.568	1.181
Cluster 4	0.168	0.926	0.568	0.000	0.984
Cluster 5	0.944	1.147	1.186	0.984	0.000

Table 18 shows the distance between cluster centroid chemical oxygen demand of clustering generated by K-Means.

Table 18

Distance between cluster centroid of total suspended solids.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster 1	0.000	0.644	0.203	0.535	0.187
Cluster 2	0.644	0.000	0.708	0.735	0.770
Cluster 3	0.203	0.708	0.000	0.691	0.137
Cluster 4	0.535	0.735	0.691	0.000	0.697
Cluster 5	0.187	0.770	0.137	0.697	0.000

Table 19 shows the distance between cluster centroid chemical oxygen demand of clustering generated by K-Means.

Table 19

Distance between cluster centroid of dissolved oxygen.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster 1	0.000	1.069	1.281	0.894	0.971
Cluster 2	1.069	0.000	1.234	1.007	1.155
Cluster 3	1.281	1.234	0.000	0.827	0.957
Cluster 4	0.894	1.007	0.827	0.000	0.187
Cluster 5	0.971	1.155	0.957	0.187	0.000

Figure 2 shows the K-means cluster with adjusted distance between cluster centroid of 33 sampling site in Surabaya river with five cluster becomes the best alternative cluster. Figure 3 shows the Silhouette plot for 5 cluster from Davies Bouldin Index is 53.742. In this situation, the optimal number of cluster produced 672 iteration

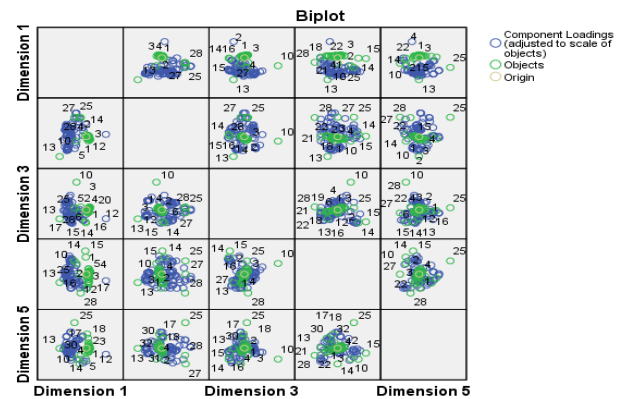


Fig. 2. Biplot distance between cluster centroid for 5 cluster.

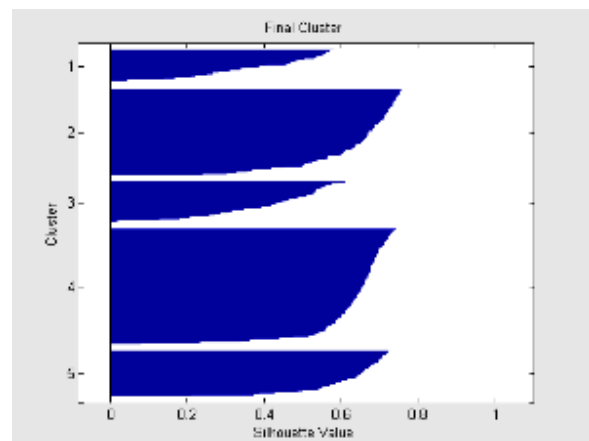


Fig. 3. Silhouette plot for 5 cluster.

5. Conclusion

From the process clustering of K-means for cluster a water quality Surabaya, it's get index value Davies Bouldin minimal 53.742 with learning rate 0.005 with 672 iteration. The result of K-means description to the pollution in Surabaya from industrial sector which are in around the Surabaya river.

Environmental issues, especially water quality in Surabaya as a source for drinking water Surabaya community, can be done by clustering the data quality of the water. Input to the K-means is influenced by water quality parameters as attributes of training, clusters are formed, learning rate is applied, and the number of epochs on the network. The resulting cluster gives an overview of the distribution of pollutants in Surabaya which includes parameter biological oxygen demand, chemical oxygen demand, total suspended solid, and dissolved oxygen. The results of the K-means cluster training is beneficial for consideration in future decisions. Especially in liquid waste problem in Surabaya at the point of the resulting cluster.

Acknowledgments

This purpose research support from the Institut Teknologi Sepuluh Nopember Surabaya in Indonesia, Faculty of Mathematics and Natural Science Department of Mathematics, Environmental Engineering Faculty of Civil Engineering and Planning Department of Environmental Engineering, and the Perum Jasa Tirta I Malang are highly appreciated.

References

- Ay, M., Kisi, O. 2012. Modeling of dissolved oxygen concentration using different neural network techniques in Foundation Creek, El Paso County, Colorado, USA. *J. Environ. Eng.* 138 (6), 654–662.
- Davies, David L.; Bouldin, Donald W. 1979. "A Cluster Separation Measure". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. PAMI-1 (2): 224–227.
- Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136, A K-means clustering algorithm. *J. Roy. Statist. Soc., Ser. C (Applied Statistics)* 28 (1), 100–108.
- Irawan, M.I., Syaharuddin, Daryono B.U., Alvida M.R. 2013, Intelligent Irrigation Water Requirement System Based on Artificial Neural Networks and Profit Optimization for Planning Time Decision Making of Crops in Lombok Island. *Journal of Theoretical and Applied Information Technology* 31st December 2013. Vol. 58 No. 3
- Jie, L., Peng, Y., Wensheng, L., Agudamu, L., 2014. Comprehensive Assessment Grade of Air Pollutants based on Human Health Risk and ANN Method. *Procedia Eng.*, 2014 International Symposium on Safety Science and Technology 84, 715–720. doi:10.1016/j.proeng.2014.10.483
- Li, Y., Wu, H., 2012. A Clustering Method Based on K-Means Algorithm. *Phys. Procedia*, International Conference on Solid State Devices and Materials Science, April 1-2, 2012, Macao 25, 1104–1109. doi:10.1016/j.phpro.2012.03.206
- Milligan, G.W., Cooper, M.C., 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50, 159–179.
- Nurul, E. H., Karnaningroem, N., 2012. The implementation of hydrodynamic model in water treatment to estimate turbidity removal. *International Journal of Environmental Science and Research* Vol.2, No.1, 2012, pp.129-133
- Rahmawati, S.F., M. Isa Irawan, Nieke Karnaningroem, 2014a, Application of Kohonen Self Organizing Maps for Clustering Quality of Water Surabaya river. Working Paper. Faculty of Mathematics and Natural Science, Universitas Negeri Semarang, November, 8st.
- Rahmawati, S.F., M. Isa Irawan, Nieke Karnaningroem, 2014b, Distribution of Pollutant in Surabaya River Using Kohonen Neural Network. Working Paper. Faculty of Environmental, Institut Teknologi Sepuluh Nopember Surabaya, December 3st.
- Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Comput. Appl. Math.* 20, 53–65.
- Singh, K.P., Basant, A., Malik, A., Jain, G., 2009. Artificial neural network modeling the river water quality – a case study. *Ecol. Model.* 220, 888–895
- Su, M.R., Yang, Z.F., Chen, B., 2012. Ecosystem health assessment of urban clusters: Method and application. *Procedia Environ. Sci.*, 18th Biennial ISEM Conference on Ecological Modelling for Global Change and Coupled Human and Natural System 13, 1134–1142. doi:10.1016/j.proenv.2012.01.106
- Surabaya, 2014. . Wikipedia Free Encycl.
- Zeľeňáková, M., Čarnogurská, M., Gargar, I., 2012. Prediction of Pollutant Concentration in Laborec River Station, Slovak Republic. *Procedia Eng.*, CHISA 2012 42, 1474–1483. doi:10.1016/j.proeng.2012.07.540
- Zhang, Y., Yang, Z., 2007. Eco-efficiency of urban material metabolism: a case study in Shenzhen, China. *Acta Ecol. Sin.* 27, 3124–3131. doi:10.1016/S1872-2032(07)60067-5
- Zhao, Y., Xia, X.H., Yang, Z.F., Wang, F., 2012. Assessment of water quality in Baiyangdian Lake using multivariate statistical techniques. *Procedia Environ. Sci.*, 18th Biennial ISEM Conference on Ecological Modelling for Global Change and Coupled Human and Natural System 13, 1213–1226. doi:10.1016/j.proenv.2012.01.115
- Zirnea, S., Lazar, I., Foudjo, B.U.S., Vasilache, T., Lazar, G., 2013. Cluster Analysis Based of Geochemical Properties of Phosphogypsum Dump Located Near Bacau City in Romania. *APCBEE Procedia*, 4th International Conference on Environmental Science and Development- ICESD 2013 5, 317–322. doi:10.1016/j.apcb.2013.05.054

Author Biographies

Sri Rahmawati Fitriatien, master student of Mathematics Faculty at Institut Teknologi Sepuluh Nopember (ITS) Surabaya. Currently undergoing final semester in the Department of Mathematics. The author has conducted

researchh “Application of Kohonen Self Organizing Maps for Clustering Quality of Water Surabaya river” and working paper at Faculty of Mathematics and Natural Science, Universitas Negeri Semarang, 8 November 2014. And the author has research “Distribution of Pollutant in Surabaya River Using Kohonen Neural Network” and working paper at Faculty of Environmental, Institut Teknologi Sepuluh Nopember Surabaya, 3 December 2014.

Mohammad Isa Irawan, Professor of Mathematics and Computer Sciences from Institut Teknologi Sepuluh Nopember (ITS) Surabaya. The author published articles areas of interest Artificial Intelligence, Bioinformatics, and Decision Support System. Citation indices since 2009 until now and several paper were published already indexed Scopus. Among the paper that have been published in 2013 is “Intelligent Irrigation Water Requirement System Based on Artificial Neural Networks and Profit Optimization for Planting Time Desicion Making of Crops in Lombok Island”, Journal of The Theoretical and Applied Information Technology 31st December 2013. Vol. 58 No.3. The ohter published is “ The Implentation of Multi-Attribute Approach in Decision Making foe Defense Sea Region Models”, Journal of The Theoretical & Apllied Information Technology, Volume 53, 2013. And “Implementation of Nondominated Sorting Genetic Algorithm-II (NSGA-II) for Multiobjective Optimization Problems on Distribution of Indonesian Navy Warship”, Journal of Theoretical & Applied Information Technology 64 (1), 2014. And “The Implemtation of Binary Genetic Algorithm (BGA) for Optimizing the Task of Indonesian Navy Ship Patrols Related to the Security of Indonesian Seas”, Journal of Theoretical & Apllied Information Technology 67 (1), 2014

Nieke Karnaningroem, Professor of Environmental Quality Management, Modelling of Environmental from Institut Teknologi Sepuluh Nopember (ITS) Surabaya. The author citation indices since 2009 until now and several paper were published already indexed Scopus. Among the paper that have been published is “The Implementation of Hydrodynamics Model in Water Treatment to Estimate Turbidity Removal”, International Journal of Environmental Science and Research, Vol. 2, No.1, 2012, pp.129-133. The others published is “A Review in Increasing Biodegradability of Wastewater for Biological Process”, International Journal of Academic Research, Volume 2 (5), 2010. And “Model of Hydrodynamics for Settling Floes in Rectangular Sedimentation Basin”, Journal of Mathematics and Technology (3), 2010.