An International Journal
JSCDSS
http://www.jscdss.com
Vol.1 No.1 December 2014: 30-38
Article history:
Accepted 16 December 2014
Published online 16 December 2014

Journal of Soft Computing and Decision Support Systems

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 desceibes the shape of the distibution of pollutants proximity region gruop. 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). Surabay 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 cllassification by considering physical, chemical, bilogical 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 impportant 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 forcesting (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 classifiying 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 km^2, consisting of 289.7 km^2 watershed Kali Marmoyo,watershed area of 99.4 km^2 Kali Watudakon, and some of the tributary watershed covering an area 227.3 km^2 (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 metdhod (Zirnea et al., 2013).

K-means clustering technique is one of the clustering techniques and has been used in many areas. This tecniques is based on distance matrix (Hartigan and Wong, 1979). Euclidean distance is used as s distance criterion. Algorithm starts with $\boldsymbol{k}$ initial seeds of clustering. All $\boldsymbol{n}$ 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 becames minimum value. The accuracy of the k-means procedure is very dependent on the choice of the initial seed (Milligan and Cooper, 1985).

Silhoutte 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; $\boldsymbol{J_i}$ is the Euclidean distance between $\boldsymbol{i}$ data vector $(\boldsymbol{x_k})$ and $\boldsymbol{j}$ clsuetr centre $(\boldsymbol{c_i})$; $\boldsymbol{c}$ is the number of cluster $(\boldsymbol{c} \in \boldsymbol{N} \text{ and } \boldsymbol{2} \le \boldsymbol{c} \le \boldsymbol{n})$, and $\boldsymbol{u_{ij}}$ is $\boldsymbol{i}$ membership degree. These parameters are calculated with Eqs. (4) and (5) according to the conditions (3).

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

$$\boldsymbol{U} = \begin{bmatrix} u_{11} & . & u_{1n} \\ . & . & . \\ . & . & . \\ u_{c1} & . & u_{cn} \end{bmatrix}_{cxn} \tag{2}$$

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

$$\boldsymbol{u_{ij}} = \begin{Bmatrix} 1 \; if \; \|x_j - c_i\|^2 \le \|x_k - c_i\|^2 \\ 0 \; if \; \|x_j - c_i\|^2 \le \|x_k - c_i\|^2 \end{Bmatrix}; \text{for } k \ne i \tag{3}$$

$$|\boldsymbol{G_i}| = \sum_{j=1}^{n} \boldsymbol{u_{ij}} \tag{4}$$

$$\boldsymbol{c_i} = \frac{1}{|G_i|} \sum_{k,x_k \in G_i} \boldsymbol{x_k} \tag{5}$$

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).

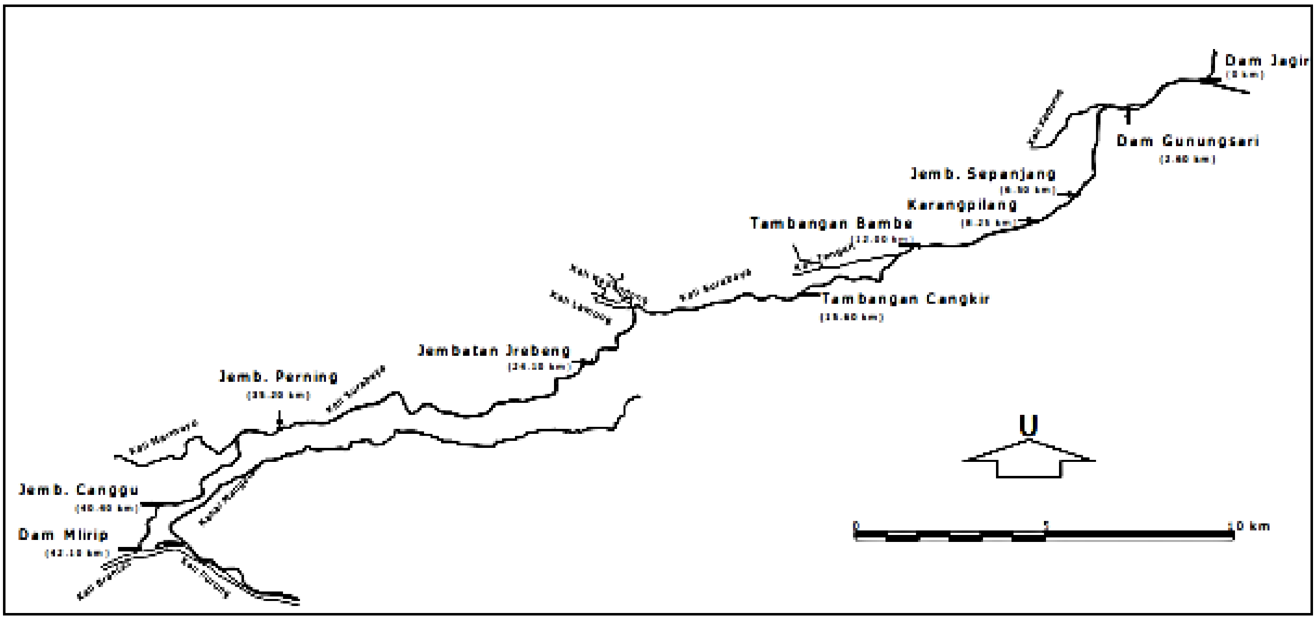


Fig. 1. Location and sampling sites in Surabaya river

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 cluctering 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.

Th 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
Custer 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
Custer 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
Custer 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
Custer 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	0.569	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 suspend 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 Silhoutte plot for 5 cluster from Davies Bouldin Index is 53.742. In this situation, the optimal number of cluster produced 672 iteration

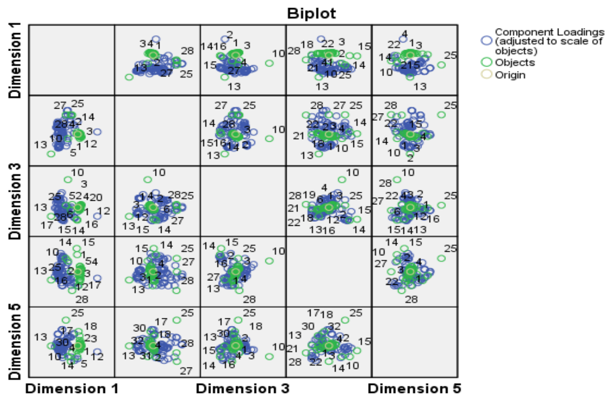


Fig. 2. Biplot distance between cluster centroid for 5 cluster.

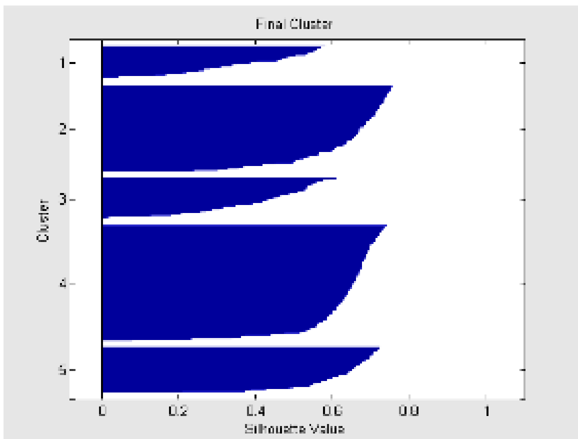


Fig. 3. Silhoutte plot for 5 cluster.

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 suspend 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 higly 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 Kournal 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.

Zeleňáková, 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.apcbee.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 researdh "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 & Apllied 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 Flocs in Rectangular Sedimentation Basin", Journal of Mathematics and Technology (3), 2010.

Prosiding Seminar Teknologi Lingkungan 2014 : 3 Desember 2014
ISBN 978-602-95595-9-0

Pola Sebaran Polutan di Kali Surabaya Menggunakan Jaringan Kohonen

Sri Rahmawati F. [1], M. Isa Irawan [2], Nieke Karnaningroem [3]

Mahasiswa Program Magister, Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam,Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia [1]*
amhar.be@gmail.com
Profesor, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia [2]
Profesor, Jurusan Teknil Lingkungan, Fakultas Teknik Sipil dan Perencanaan, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia [3]

Abstrak

Objek penelitian ini adalah Kali Surabaya yang menjadi sumber air baku air minum masyarakat Surabaya. Aliran sungai dari Kali Surabaya dimulai dari DAM Mlirip (Mojokerto) sebagai hulu sungai melewati Sidoarjo, Gresik, dan sampai pada hilir sungai yaitu DAM Jagir Wonokromo (Surabaya). Latar belakang kajian penelitian karena air permukaan Kali Surabaya mengalami penurunan kualitas air yang dirasakan semakin hari semakin meningkat akibat sebagian besar limbah industri cair dari kegiatan manusia dibuang ke saluran yang bermuara di Kali Surabaya baik secara langsung ataupun tidak. Komponen pencemar air Kali Surabaya adalah kebutuhan oksigen biokimiawi yang dikenal dengan BOD dan kebutuhan oksigen kimia yang dikenal dengan COD. Bahkan untuk komponen BOD di beberapa titik pantau berstatus cemar berat dengan konsentrasinya melebihi baku mutu air kelas II. Jaringan Kohonen merupakan metode untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimiliki. Objek kajian pada penelitian ini adalah *point source* yang pembuangan limbah cair industri bermuara di Kali Surabaya sejak tahun 2011 hingga 2013. Langkah pertama kajian penelitian ini adalah normalisasi data kualitas air BOD dan COD Kali Surabaya. Hasil dari langkah ini adalah nilai konsentrasi BOD dan COD yang berada dalam rentang 0 dan 1. Nilai konsentrasi yang ternormalisasi diterapkan pada Jaringan Kohonen sehingga menghasilkan model dalam bentuk sebaran polutan yang menggambarkan kedekatan kelompok wilayah. Dari hasil pelatihan Jaringan Kohonen diperoleh kesesuaian kualitas air setiap wilayah di sepanjang aliran Kali Surabaya yang terbentuk dari kesamaan dan kemiripan atribut. Pola sebaran polutan yang di dapatkan adalah *cluster* wilayah tinggi BOD, rendah COD, dan tinggi BOD dan COD.

Kata Kunci : Jaringan Kohonen, Kualitas Air, Kali Surabaya, Sebaran Polutan

1. Pendahuluan

Kali Surabaya merupakan salah satu air permukaan yang digunakan oleh masyarakat Surabaya untuk memenuhi kebutuhan sehari-hari. Kali Surabaya memiliki luas daerah aliran sungai (DAS) 630,7 km^2. Salah satu pemanfaatannya sekitar 96% air baku PDAM Kota Surabaya dipasok dari Kali Surabaya. Kali Surabaya merupakan sungai terpanjang di Kota Surabaya dengan panjang 17.400 meter sehingga keberadaan sungai di Surabaya sebagai sumber air bagi kelangsungan hidup masyarakat memliki arti yang sangat penting, termasuk Kali Surabaya. Berdasarkan pemantauan Badan Lingkungan Hidup Kota Surabaya tahun 2013, , status mutu air Kali Surabaya selama satu tahun pemantauan menunjukkan 69,45% berstatus cemar ringan, 22,22% berstatus cemar sedang, dan 8,33% berstatus cemar berat dengan parameter BOD dan TSS konsentrasinya melebihi baku mutu air kelas II [1]. Dengan kata lain, air Kali Surabaya tidak layak untuk dijadikan sebagai air baku untuk memenuhi kebutuhan air minum manusia . Sementara itu, bila ditinjau dari segi kuantitas, Perum Jasa Tirta telah memprediksi bahwa pada tahun 2025, Surabaya akan mengalami defisit air bersih sebesar 7,43 m^3/detik [6].

Masalah utama Kali Surabaya adalah besarnya kandungan limbah cair hasil dari kegiatan manusia yang dibuang ke aliran Kali Surabaya. Limbah tersebut berasal dari pemukiman, industri, pertanian, peternakan, dan lain-lain. Pembangunan di Kota Surabaya khususnya pada sektor industri telah membawa dampak bagi kehidupan manusia, baik dampak positif maupun negatif. Salah satu dampak negatif yang ditimbulkan adalah pencemaran air yang dirasakan

ISBN 978-602-95595-9-0

semakin meningkat. Dengan meningkatnya pencemaran air di Kali Surabaya, diperlukan pengelompokan *point source* berdasarkan beban polutan, sehingga dapat diketahui sebaran polutan di sepanjang Kali Surabaya sebagai informasi dalam kebijaksanaan pembangunan Kota Surabaya. Penelitian ini bertujuan untuk mengimplementasikan algoritma Jaringan Kohonen dalam data *clustering* data kualitas air Kali Surabaya tahun 2013 sektor industri sehingga mendapatkan karakteristik data dari hasil pelatihan Jaringan Kohonen.

2. Metode yang Diterapkan

2.1. Pembersihan Data

Data yang bersih adalah data yang tidak memiliki nilai yang tidak lengkap dan *noise*. Proses pembersihan data bertujuan untuk melengkapi nilai yang tidak lengkap, memperhalus *noise* ketika teridentifikasi, dan memperbaiki ketidakkonsistenan data [7]. Setelah dilakukan proses pembersihan data, dilakukan *preprocessing* data dengan menggunakan *Principal Component Analysis* (PCA) guna memperoleh komponen yang menjadi variabel bebas baru yang dianalisa pengeruhnya terhadap variabel tak bebas. Variabel bebas baru dalam penelitian ini parameter kualitas air adalah *Total Suspended Solid* (TSS) sedangkan variabel tak bebas adalah *Biochemiycal Oxygen Demand* (BOD), *Chemical Oxygen Demand* (COD) dan *Dissolved Oxygen* (DO).

2.2 Jaringan Syaraf Tiruan

Jaringan syaraf merupakan representasi dari otak manusia yang mensimulasikan proses pembelajaran pada otak manusia. Penggunaan istilah jaringan syaraf tiruan karena jaringan syaraf ini diimplementasikan dengan menggunakan program komputer yang mampu menyelesaikan sejumlah proses perhitungan selama proses pembelajaran. Jaringan syaraf tiruan diperkenalkan oleh ilmuwan Finlandia bernama Teuvo Kohonen pada tahun 1982.

Jaringan Syaraf Tiruan (JST) merupakan suatu model komputasi yang *powerfull* yang dapat membuat model suatu fungsi yang sangat kompleks dengan menyelesaikan masalah yang membutuhkan *input* untuk mendapatkan *output* dari permasalahan. Jaringan syaraf tiruan menirukan cara kerja sistem otak manusia yaitu jaringan syaraf biologi sehingga memiliki kemampuan untuk belajar dan beradaptasi terhadap *inputan-inputan* yang diberikan.

Jaringan syaraf tiruan telah banyak diaplikan ke dalam *real condition* dan berbagai disiplin ilmu. Sesuatu permasalahan yang menggunakan Jaringan Syaraf Tiruan dapat diprediksi, dikelompokkan dan dikontrol. Salah satunya adalah untuk memprediksi curah hujan, kelembapan, kecepatan angin, dan sinar matahari dengan tingkat akurasi rata-rata mencapai 99,76%. Jaringan syaraf tiruan digunakan untuk memperkirakan perencanaan pola tanam tahun 2012 di wilayah Pulau Lombok. Dengan penerapan jaringan syaraf tiruan pada pola tanam, petani mendapatkan keuntungan rata-rata meningkat sebesar 12.10% dari tahun sebelumnya [8]. Dalam ilmu kedokteran, jaringan syaraf tiruan dapat digunakan untuk memprediksi *cervical cancer*. Model jaringan syaraf tiruan yang digunakan untuk deteksi *cervical cancer* ini memiliki kualitas yang baik yaitu persentase sensibility sebesar 98% dengan spesifikasi 97%. Hal ini menunjukkan bahwa dengan jaringan syaraf tiruan memberikan hasil yang baik karena memiliki kemampuan yang kuat untuk mencocokkan dan meramalkan kondisi yang akan datang [9].

2.3 Jaringan Kohonen

Jaringan Kohonen merupakan metode berdasarkan model dari pendekatan jaringan syaraf tiruan. Metode pembelajaran pada Jaringan Kohonen adalah *unsupervised* yaitu belum diketahui *output* dari proses pelatihan jaringan [5]. Jaringan Kohonen terdiri dari dua *layer* yaitu *input layer* dan *output layer*. Setiap *neuron* pada *input layer* akan terhubung dengan setiap *neuron* pada *output layer*. Setiap *neuron* dalam *output layer* merepresentasikan kelas dari *input* yang diberikan. Gambar struktur Jaringan Kohonen terlihat seperti pada Gambar 1. Pemanfaatan Jaringan Kohonen digunakan oleh Olawoyin pada tahun 2013 untuk mengenali pola spacial di zona yang terkontaminasi pencemar air dengan mengidentifikasi sumber pencemar kualitas air [10].

Tahapan awal dari penelitian ini adalah dilakukan normalisasi data sehingga data berada pada selang [0,1] untuk tujuan pelatihan jaringan. Data yang telah diolah kemudian dilakukan proses

ISBN 978-602-95595-9-0

kerja dengan dilakukan proses *training* dan *testing* data menggunakan Jaringan Kohonen. Untuk normalisasi data digunakan rumus sebagai berikut [3] :

$$f(x) = \frac{x - x_{min}}{x_{maks} - x_{min}}$$

Kemampuan Jaringan Kohonen adalah memecahkan permasalahan yang spesifik karena memiliki struktur sistem pengolahan informasi. Kemampuan ini didukung karena adanya *neuron-neuron* yang saling bekerja saling berhubungan antara *input* dengan *output* yang akan dihasilkan dari sistem yang bekerja [4]. Kemampuan ini memetakan pola *inputan* melalui model pembelajaran *unsupervised*, dimana hasil pemetaan tersebut menunjukkan adanya keterkaitan di antara pola *inputan* dengan mempertahankan hubungan topologi [2]. Pemanfaatan Jaringan Kohonen dalam pembuatan peta masukan dapat membagi *inputan* menjadi beberapa *cluster* yang memiliki kemiripan antara *inputan* objek satu dengan *inputan* objek yang lainnya.

Pada Jaringan Kohonen, *neuron* bertugas sebagai penerima informasi untuk diteruskan pada *soma* (badan sel) untuk dilakukan proses pengolahan informasi, kemudian diteruskan pada *neurit* (neurit) yang menghubungkan ke sel syaraf lainnya dengan tujuan mengirimkan *impuls* [5].

Data yang digunakan dalam penelitian ini adalah data sekunder kualitas air Kali Surabaya dari Perum Jasa Tirta I Malang tahun 2013. Titik pantau yang diamati adalah industri yang pembuangan limbah cair bermuara di Kali Surabaya. Kawasan industri merupakan salah satu *point source* yang menyebabkan penurunan kualitas air Kali Surabaya. Berikut akan disajikan beberapa industri yang berada di aliran Kali Surabaya :

1.	Saluran Limbah Pulo Wonokromo (*IC* 1)	2.	Saluran Ketintang (*IC* 2)
3.	Pakuwon (*IC* 3)	4.	Perusahaan Tahu Halim (*IC* 4)
5.	Perusahaan Tahu Gunungsari (*IC* 5)	6.	Pemotongan Hewan KMS (*IC* 6)
7.	PT. Gawerejo (*IC* 7)	8.	PT. Suparma (*IC* 8)
9.	Wandira Kencana Alam (*IC* 9)	10.	Spindo (*IC* 10)
11.	Pabrik Tahu Purnomo (*IC* 11)	12.	Wiramas Init Lestari (*IC* 12)
13.	PT. Aneka Multi Pangan (*IC* 13)	14.	PT. Timur Megah Steel (*IC* 14)
15.	PT. Tirta Utama Raharjo (*IC* 15)	16.	PT. Titani Alam Semesta (*IC* 16)
17.	PT. Hueychyi (*IC* 17)	18.	PT. Miwon Indonesia (*IC* 18)
19.	Accasia Nusantara Indah (*IC* 19)	20.	Sosro Kencono (*IC* 20)
21.	PT. Pajawali Bolang (*IC* 21)	22.	PT. Spindo Makmur (*IC* 22)
23.	PT. Wood Indonesia (*IC* 23)	24.	Perusahaan Tahu Sidomakmur (*IC* 24)
25.	Pabrik Acithelyn (*IC* 25)		

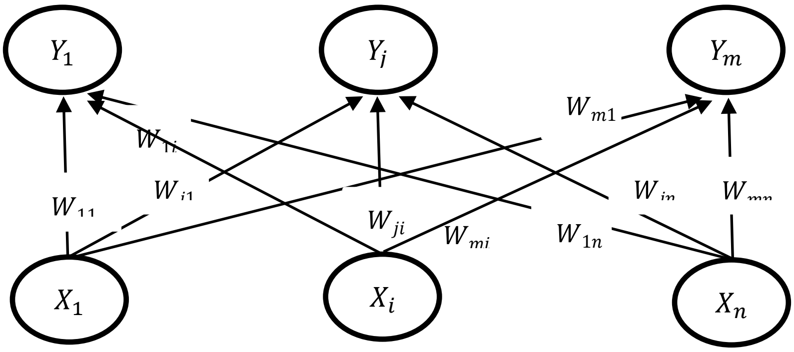
Sedangkan atribut-atribut yang digunakan dalam penelitian ini terdiri dari :

x_1 = *Biochemiycal Oxygen Demand* (BOD)
x_2 = *Chemical Oxygen Demand* (COD)
x_3 = *Total Suspended Solid* (TSS)
x_4 = *Dissolved Oxygen* (DO)

2.3 Algoritma Jaringan Kohonen

Misalkan himpunan dari *m* adalah nilai-nilai untuk *record* ke-*n* menjadi sebuah vektor input $x_n = x_{n1}, x_{n2}, \dots, x_{nm}$, dan himpunan dari *m* bobot untuk simpul *output* tertentu *j* menjadi vektor bobot $w_j = w_{j1}, w_{j2}, \dots, w_{jm}$ [5].

ISBN 978-602-95595-9-0



Gambar 2.1 Arsitektur Jaringan Kohonen

Algoritma pada jaringan Kohonen diterapkan dalam pengelompokan data adalah sebagai berikut [5]:

Langkah 0 Inisialisasi pembobotan w_{ij} dengan random. Menset parameter *learning rate* (α), parameter radius *neightbourhood* (R).
Langkah 1 Apabila kondisi selesai belum terpenuhi, lakukan langkah 2-8
Langkah 2 Untuk tiap vektor input *x* (x_i, *i* =1,…,n), lakukan langkah 3-5
Langkah 3 Untuk tiap *j* (*j* =1,...m), hitung jarak Euclidean

$$D(j) = \sum_i (w_{ij} - x_i)^2$$

Langkah 4 Mencari indeks *j* dengan jarak *D*(*j*) terdekat (minimum)
Langkah 5 Melakukan perbaikan nilai w_{ij} dengan nilai tertentu, yaitu:

$$w_{ij}(new) = w_{ij}(old) + \alpha\left[x_i - w_{ij}(old)\right]$$

Langkah 6 Melakukan update learning rate

$$\alpha\ (new) = \ a.\alpha(old)$$

Langkah 7 Mereduksi radius dari fungsi tetangga pada waktu tertentu (*epoch*)
Langkah 8 Uji kondisi penghentian.
Proses pembelajaran akan berlangsung terus hingga mencapai maksimum *epoch*.

2.4 Validasi *Cluster*

Proses pelataihan Jaringan Kohonen akan menghasilkan *cluster* yang bersesuain berdasarkan kemiripan karakteristik dari objek pengamatan. Pada penelitian ini, *cluster* yang dihasilkan akan divalidasi dengan menggunakan Indeks Davies-Boudin.

Pendekatan pengukuran ini untuk memaksimalkan jarak *inter-cluster* di antara *Cluster* C_i dan C_j dan pada saat yang bersamaan meminimalkan jarak antara titik dalam sebuah *cluster*. Jarak *intra-cluster* $S_c(Q_k)$ dalam *Cluster* Q_k adalah [7]

$$S_c(Q_k) = \frac{\sum_i \|X_i - C_k\|}{N_k}$$

dengan C_k adalah *centroid* dari *Cluster* Q_k sedangkan N_k adalah banyak titik yang termasuk dalam *Cluster* Q_k. Jarak *inter-cluster* didefinisikan sebagai

$$d_{kl} = \|C_k - C_l\|$$

dengan C_k dan C_l adalah *Centroid Cluster k* dan *Centroid Cluster l.* Indeks Davies- Bouldin didefinisikan sebagai berikut

$$DB\ (nc) = \frac{1}{n}\sum_{k=1}^{nc} max\left\{\frac{S_c(Q_k) + S_c(Q_l)}{d_{kl}\ (Q_k, Q_l)}\right\}$$

dengan *nc* adalah banyak *cluster*. Suatu *cluster* dikatakan optimal jika memiliki Indeks Davies-Bouldin minimal.

ISBN 978-602-95595-9-0

3. Pembahasan Hasil

Validitas hasil *cluster* yang dilakukan adalah memperoleh Indeks Davies-Bouldin yang minimum dari proses pelatihan Jaringan Kohonen. Inisialisasi awal yang dilakukan pada Jaringan Kohonen adalah penentuan parameter-parameter meliputi *learning rate* yang digunakan, penurunan *learning rate*, *radius neighbourhood* akan mempengaruhi hasil pelatihan jaringan.

Nilai Indeks Davies-Bouldin yang dihasilkan berdasarkan *inisialisasi* awal yang digunakan memiliki ukuran vektor bobot yang berbeda tampak pada Tabel 3.1

Tabel 3.1 Nilai Indeks Davies-Bouldin

Ukuran *Output*	***Learning Rate***	**Iterasi**	**Indeks Davies-Bouldin**
3	0.1	209	209.285
4	0.05	139	353.452
5	0.15	194	202.856
6	0.2	374	164.302
7	0.25	263	113.370
8	0.3	773	87.917
9	0.4	672	53.742
10	0.35	536	79.743

Dari hasil pelatihan, Indeks Davies-Bouldin terbaik dihasilkan dengan parameter awal untuk ukuran *output* 9, dengan *learning rate* 0.4, iterasi sebanyak 672 iterasi, dan nilai Indeks Davies-Bouldin yang dihasilkan adalah 53.742. Banyaknya data kualitas air untuk masing-masing *cluster* dengan ukuran *output* 9 dapat ditunjukkan pada Tabel 2. Sedangkan untuk rataan dan *centroid* masing-masing *cluster* ditunjukkan pada Tabel 3.2 dan Tabel 3.3

Tabel 3.2 Banyak Anggota Masing-masing *Cluster* dengan Ukuran *Output* 9

Cluster ke-	**Banyak Anggota**	**Persentasi Banyak Anggota**
1	2	9.90
2	3	14.96
3	4	10.37
4	2	11.16
5	3	10.48
6	2	12.80
7	3	14.32
8	3	9.95
9	3	6.06

Tabel 3.3 *Centroid* Masing-masing *Cluster* dengan Ukuran *Output* 9

Cluster Ke-	***Centroid***			
	BOD	**COD**	**TSS**	**DO**
1	69.62	69.76	65.20	66.34
2	68.49	67.43	68.89	69.51
3	83.39	80.48	81.34	79.50
4	74.15	76.03	74.74	71.52
5	74.77	68.84	76.28	76.33
6	77.12	70.97	70.43	75.55
7	75.84	67.27	69.41	66.65
8	73.38	68.79	76.81	68.22

ISBN 978-602-95595-9-0

9	73.18	77.11	79.35	82.17

Tabel 3.4 *Mean* Masing-masing *Cluster* dengan Ukuran *Output* 9

Cluster Ke-	***Mean***				
	BOD	**COD**	**TSS**	**DO**	***Mean***
3	83.31	80.37	81.13	79.66	81.12
9	73.33	76.86	78.61	81.20	77.50
5	74.99	68.83	76.57	76.26	74.16
4	74.11	75.87	74.61	71.73	74.08
6	76.28	70.74	69.92	75.41	73.09
8	73.47	68.82	76.94	68.42	71.91
7	76.12	67.44	69.68	66.94	70.04
2	68.80	67.39	68.88	69.50	68.59
1	69.78	69.88	65.46	66.60	67.93
Mean	74.44	71.80	73.53	72.86	73.16

Cluster 3 yang memiliki 10.37% dari data (Tabel 2) adalah *cluster* yang memiliki rataan BOD, COD, TSS dan DO tertinggi (Tabel 3.4). Tetapi *cluster* 3 bukan *cluster* yang memiliki nilai Indeks Davies-Bouldin yang terbaik dari seluruh atribut.

Cluster 9 yang memiliki 6.06% dari data (Tabel 3.2), menduduki peringkat ke dua dari rataan secara keseluruhan (Tabel 4). *Cluster* 9 memiliki penyebaran polutan yang cukup banyak untuk parameter DO, TSS dan COD, tetapi sedikit penyebaran polutan untuk parameter BOD.

Cluster 5 (10.48% dari data) adalah *cluster* yang menduduki peringkat ketiga dari rataan (Tabel 2 dan Tabe 4). *Cluster* 5 memiliki pola sebaran yang sedikit untuk parameter COD dengan *mean* COD *cluster* 5 dibawah rata-rata, yaitu 68.83% dari rata-rata 71.80% (Tabel 3.4).

Cluster 4 (11.16% dari data) adalah *cluster* yang menduduki peringkat ke empat dari *mean* keseluruhan (Tabel 3.2 dan Tabel 3.4).

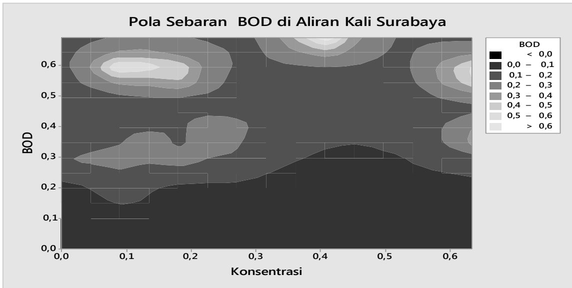
Cluster 6 menempati peringkat ke 5 untuk *mean* keseluruhan (Tabel 4), memiliki anggota terbanyak ke 2 dari data yaitu 12.80% (Tabel 3.2).

Cluster 8 yang menduduki peringkat ke enam memiliki 9.95% dari data (Tabel 2 dan Tabel 4). Secara keseluruhan, *mean* dari parameter kualitas air untuk *cluster* 8 berada di bawah rata-rata yaitu 71.91 dari rata-rata 73.16 (Tabel 3.4).

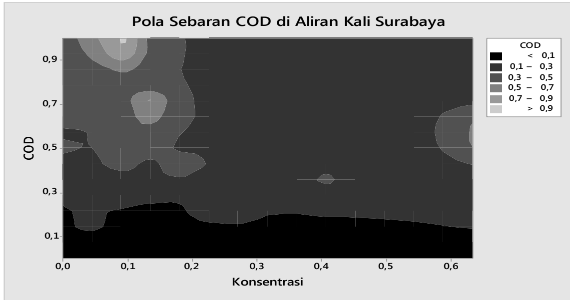
Cluster 7, 2, dan 1 merupakan 3 *cluster* yang memiliki nilai rata-rata dibawah *mean* keseluruhan.

Berdasarkan hasil uji pelatihan beban pencemaran air di Kali Surabaya terjadi kemiripan antar satu dengan yang lain yang bergabung membentuk *cluster*. *Input* data yang sudah ternormalisasi kemudian dilakukan pelatihan jaringan dengan iterasi maksimal yaitu 1000 *epoch*. Hasil pelatihan jaringan dari diperoleh keluaran sebaran polutan di Kali Surabaya dari sektor industri sebagai berikut :

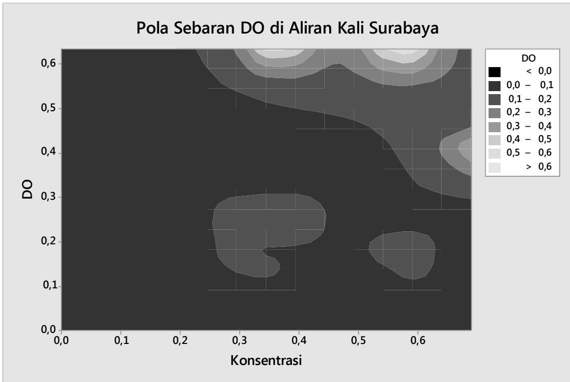
ISBN 978-602-95595-9-0



Gambar 3.1 Sebaran BOD



Gambar 3.2 Sebaran COD



Gambar 3.3 Sebara

Proses simulasi dengan Jaringan Kohonen untuk ketiga atribut parameter kualitas air Kali Surabaya, diperoleh *clustering* sumber pencemar dari sektor industri di Kali Surabaya menjelaskan jenis industri yang melakukan proses produksi denga membuang limbah cair ke aliran muara Kali Surabaya.

Dari hasil *clustering* yang diperoleh di atas, terlihat bahwa banyaknya atribut yang sama akan menggambarkan kedekatan kelompok wilayah. Setelah dilakukan pengelompokan wilayah dengan Jaringan Kohonen, mampu memperoleh *cluster* sebagai sumber pencemar air kali Surabaya dari sektor industri.

4. Kesimpulan

Dari proses pelatihan Jaringan Kohonen untuk *cluster* kualitas air Kali Surabaya diperoleh nilai Indeks Davies Bouldin minimal seberar 53.742 dengan *learning rate* 0.4 dengan banyaknya iterasi 672. Hasil pelatihan Jaringan Kohonen memberikan gambaran terhadap pencemaran air di Kali Surabaya dari sektor industri yang berada di sepanjang aliran Kali Surabaya. Permasalahan lingkungan khususnya kualitas air di Kali Surabaya sebagai sumber air baku air minum masyarakat Surabaya, dapat dilakukan dengan *clustering* data kualiatas air. Inputan pada Jaringan Kohonen dipengaruhi oleh parameter kualitas air sebagai atribut pelatihan, *cluster* yang terbentuk, *learning rate* yang diterapkan, serta jumlah *epoch* pada jaringan. *Cluster* yang dihasilkan memberikan gambaran sebaran polutan di Kali Surabaya yang meliputi paraeter BOD, COD, TSS, dan DO. Hasil dari proses pelatihan Jaringan Kohonen ini bermanfaat sebagai pertimbangan pengambilan keputusan di masa depan khususnya dalam permasalahan limbah cair di Kali Surabaya pada titik *cluster* yang dihasilkan.

5. Penghargaan

Penulis mengucapkan terima kasih dan penghormatan kepada Prof. Dr. Mohammad Isa Irawan, MT dan Prof. Dr. Ir. Nieke Karnaningroem, Dipl.SE, MSc yang telah membantu keberhasilak penelitian ini. Serta penulis mengucapkan terima kasih kepada seluruh jajaran

ISBN 978-602-95595-9-0

Perum Jasa Tirta I Malang sebagai mitra dalam kajian penelitian yang dilakukan guna membenahi permasalahan lingkungan khususnya kualitas air Kali Surabaya sebagai sumber air baku air minum masyarakat Surabaya.

6. Pustaka

Anonim, (2013). Status Lingkungan Hidup Kota Surabaya (SLHD). *Laporan Penelitian Badan Lingkungan Hidup Surabaya 2013*, Kota Surabaya.

Stergiou, C. And Siganos, D., (1989). *Neural Network*. New York, Pasific Northwest National Laboratory.

Irawan, M.I., (2008). *Exploratory Data Analysis dengan JST-Kohonen SOM: Struktur Tingkat Kesejahteraan Daerah Tingkat II se Jawa Timur*. Surabaya : Institut Teknologi Sepuluh Nopember Surabaya.

Arifin Z. dan M.I. Irawan., (2009). Adaptive Sensitivity Sensitivity-based Linear Learning Method Algorithms for Data Classification. *Proceeding of 5th International Conference of Mathematics, Statistics and Their Aplications, Juni 9 -11, Bukit Tinggi – West Sumatra Indonesia.*

Fauset, L., (1994). *Fundamental of Neural Networks*. New Jersey ,Prentice Hall Inc.

Kusumawardhani, D., (2010). *Valuasi Ekonomi Air Bersih di Surabaya*. UGM, Yogyakarta.

Han J, Kamber., (2001). *Data Mining: Concept and Techniques*. New Jersey, Prentice Hall Inc.

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*

Qiu, Xiaoping., Ning Tao, Yun Tan, Xinxing Wu, (2007), Constructing of the Risk Classification Model of Cervical Cancer by Artificial Neural Network. *Expert System with Applications 32 (2007) 1094 – 1099.*

Olawoyin, R., Nieto, A., Grayson, R.L., Hardisty, F., dan Oyewolw, S., (2013), Application of Artificial Neural Network (ANN)-Self-Organizing Maps (SOM) for the Categorization of Water, Soil and Sediment Quality in Petrochemical Regions. *The Expert System with Application, 40(2013),3634-3648.*