IOP Conf. Series: Materials Science and Engineering **190** (2017) 012014 doi:10.1088/1757-899X/190/1/012014

Real Time Monitoring System of Pollution Waste on Musi River Using Support Vector Machine (SVM) Method

Muhammad Fachrurrozi, Saparudin, Erwin

Informatics Engineering, Computer Science Faculty, University of Sriwijaya.

mfachrz@unsri.ac.id, saparudin@unsri.ac.id, erwin@unsri.ac.id

Abstract. Real-time Monitoring and early detection system which measures the quality standard of waste in Musi River, Palembang, Indonesia is a system for determining air and water pollution level. This system was designed in order to create an integrated monitoring system and provide real time information that can be read. It is designed to measure acidity and water turbidity polluted by industrial waste, as well as to show and provide conditional data integrated in one system. This system consists of inputting and processing the data, and giving output based on processed data. Turbidity, substances, and pH sensor is used as a detector that produce analog electrical direct current voltage (DC). Early detection system works by determining the value of the ammonia threshold, acidity, and turbidity level of water in Musi River. The results is then presented based on the level group pollution by the Support Vector Machine classification method.

1. Introduction
Industrial development and construction resulted in an increasing number of industrial pollution on the environment. Industrial pollutions consist of waste in the form of water, gas and solid. In general, this waste is harmful because most of its components composed of addictive and its chemicals are hard to be degraded [1]. These substances have a bad impact on the environment and threaten the survival of living beings, according to Environmental Protection Agency. Several developed countries had already developed technologies to detect levels of waste in water, but because these are relatively expensive to adopt in Indonesia, it should be developed independently in the country in order to save expenses on imported goods.

Musi River flows in the center of Palembang, Indonesia, and is used for sort of needs, such as transportation, industry and household. Everyday people use its water for bathing, washing, and toilet facilities either directly from the river or trough Water Management Company of Palembang (PAM), as well as exploiting the products such as fish and other biota. On the other hand, some large companies and small companies are located around Musi River. Unfortunately, one of the routine is dumping waste into the river. Therefore, a monitoring and classification of water contamination system is needed around the Musi River. This system works in real time, so it can quickly minimize the negative impact of pollution. Water testing quality is required to determine whether the water is polluted or not, from the deviation limits of pollution. Some factors that affect water quality is the degree of acidity (pH), turbidity, and ammonia substance (NH_3).

Support Vector Machine (SVM) method applies hypothetical space on machine learning in the form of linear function in a high dimensional feature space. The learning technique based on optimization theory with learning bias implementation derived from statistical learning theory [2] [3]. This method is used to classify the level of water pollution based on the parameters of the factors affecting the level of water pollution.

2. Research Problems

The main problem of this research is the absence of public real-time monitoring of water pollution in the Musi river basin. Moreover, it is important to measure the result of the application of SVM method to classify the level of water pollution of the Musi river based on factors that affect water quality.

3. Research Objectives

This research aims to develop a real-time monitoring system of water pollution in Musi River and to know the accuracy of monitoring of water pollution level classification using Support Vector Machine (SVM).

4. Literature Review

4.1. Support Vector Machine (SVM)

The aim of learning process in SVM is to get a hypothesis in the form of best hyperplane which not only minimizes the empirical risk which is the average error on the training data, but also has a good Generalization [4]. Generalization is the ability of a hypothesis to classify data that is not contained in the training data correctly. SVM is a machine learning system that works on the principle of Structural Risk Minimization (SRM).

Figure 1 describes that SVM aims to find the best hyperplane that separates two classes in the input space. The best hyperplane separation between the two classes can be found by measuring the margin hyperplane and looking for the maximum points. The margin is the distance between the hyperplane to the nearest pattern of each class. The closest pattern is called a support vector. The solid line in Figure 1 shows the best hyperplane that is located right in the middle of both classes, while the red dots and yellow are in a black circle is a support vector. Attempts to locate hyperplane is the core of the learning process on SVM.

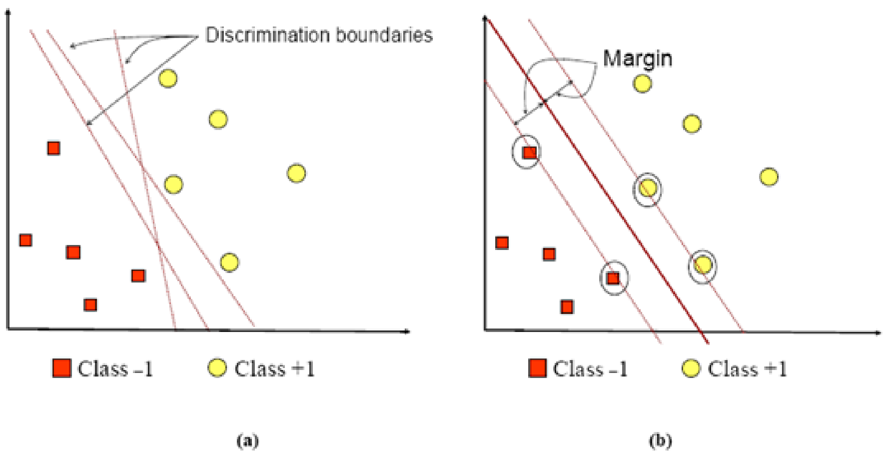

Figure 1. Class -1 and class +1 separated by Hyperplane. [5]

The provided data denoted into equation 1.

$$\vec{x_i} \in R^d \tag{1}$$

Whereas each label is denoted $Y\,i \in \{-1, +1\}$ for i = 1, 2, L, l, where *l* is the number of data. The both classes are assumed to -1 and +1 that can be completely separated by hyperplane with d dimension, which is defined in equation 2.

$$\vec{w}.\vec{x} + b = 0 \tag{2}$$

Pattern Xi which belongs to class -1 (negative samples) can be formulated as a pattern that equals to the equation 3.

$$\vec{w}.\vec{x_i} + b \leq -1 \tag{3}$$

Meanwhile x_i pattern that includes class +1 (positive samples) can be formulated as a pattern that satisfies the equation 4.

$$\vec{w}.\vec{x_i} + b \geq +1 \tag{4}$$

The maximum margin can be found by maximizing the value of the distance between the hyperplane and the closest point. It can be formulated as a Quadratic Programming (QP) problem, i.e. finding the minimum point of the equation (5), by observing the constraints in equation (6).

$$\min_{\vec{w}} \tau(w) = \frac{1}{2}\|\vec{w}\|^2 \tag{5}$$

$$y_i(\vec{x}_i.\vec{w} + b) - 1 \geq 0, \forall i \tag{6}$$

This problem can be solved with a variety of computational techniques, including Lagrange Multiplier.

$$L(\vec{w}, b, \alpha) = \frac{1}{2}\|\vec{w}\|^2 - \sum_{i=0}^{n} \alpha_i \left(y_i\left((\vec{x}_i.\vec{w} + b) - 1\right)\right) \tag{7}$$

$$(i = 1.2 \ldots, l) \tag{8}$$

α_i is Lagrange multipliers with value zero or positive ($\alpha i \geq 0$). The optimum value from equation 7, can be calculated with minimizing L against w and b, a maximize L against αi.
By assuming a characteristic that in the optimum gradient point L = 0, equation 7 can be modified as a maximizing problem that only contains αi which expressed in Equation 8.

Maximizing

$$\sum_{i=l}^{l} \alpha_i - \frac{1}{2}\sum_{i,j=l}^{l} \alpha_i \alpha_j y_i y_j \vec{x}_i.\vec{x}_j \tag{9}$$

With conditions (constraints),

$$\alpha_i \geq 0 \ (i = 1{,}2 \ldots, l) \ \sum_{i=l}^{l} \alpha_i y_i = 0 \tag{10}$$

The result of this calculation obtained some value of αi, most of them have a positive value. Data which correlated with the positive value of αi called support vector.

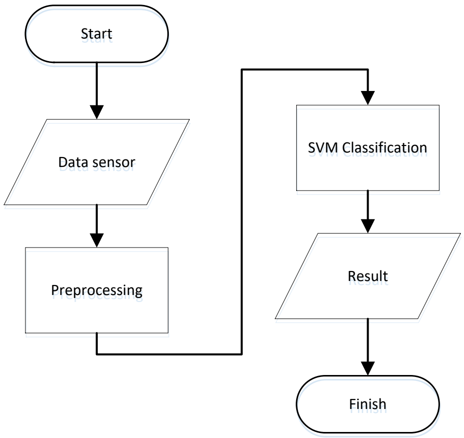

Figure 2. System Flowchart

5. Result and Discussion

Figure 2 shows a flowchart that illustrates the process of the study. Data is obtained from 3 sensors in real time, then send them to the system via GPRS, then the system display the classification result. The object research place was taken on 3 locations, they are Kalidoni, Sei Gerong, and Seberang Ulu 1 (SU 1).

Training Process

SVM method usually use training data. But in this research training process was skipped and replaced by the threshold based on Table 1.

Table 1. Parameter Threshold.

Parameter	Unit	Maximum Value
PH	Mg/l	6 - 9
NH3	Mg/l	0,5
Waste Water	Mg/l	20

Table 2. Measurement result: Seberang Ulu I

Date	**pH**	**Amonisk (%)**	**Keruh (NTU)**	**Location**	**State**
30-07-2016	4.82	3	8.48	Sebarang Ulu I	Bad
30-07-2016	4.94	2	9.02	Sebarang Ulu I	Bad
30-07-2016	5.94	2	8.44	Sebarang Ulu I	Bad
30-07-2016	5.06	2	7.09	Sebarang Ulu I	Bad
30-07-2016	4.82	1	5.68	Sebarang Ulu I	Bad
30-07-2016	4.94	2	5.98	Sebarang Ulu I	Bad
30-07-2016	4.94	2	5.92	Sebarang Ulu I	Bad
30-07-2016	5.06	2	7.27	Sebarang Ulu I	Bad
30-07-2016	5.06	2	8.19	Sebarang Ulu I	Bad
30-07-2016	4.82	2	8.95	Sebarang Ulu I	Bad
30-07-2016	5.06	2	9.58	Sebarang Ulu I	Bad
30-07-2016	5.06	3	9.73	Sebarang Ulu I	Bad
30-07-2016	5.18	3	9.79	Sebarang Ulu I	Bad

Date	pH	Amonisk (%)	Keruh (NTU)	Location	State
30-07-2016	4.94	2	8.81	Sebarang Ulu I	Bad
30-07-2016	4.82	3	8.75	Sebarang Ulu I	Bad
30-07-2016	5.06	3	8.46	Sebarang Ulu I	Bad
30-07-2016	5.06	3	9.19	Sebarang Ulu I	Bad
30-07-2016	4.94	2	8.65	Sebarang Ulu I	Bad
30-07-2016	4.94	2	8.63	Sebarang Ulu I	Bad
30-07-2016	4.94	2	7.91	Sebarang Ulu I	Bad
30-07-2016	4.94	3	7.41	Sebarang Ulu I	Bad
30-07-2016	4.94	3	7.65	Sebarang Ulu I	Bad
30-07-2016	4.94	3	7.63	Sebarang Ulu I	Bad
30-07-2016	5.06	2	7.25	Sebarang Ulu I	Bad
30-07-2016	5.06	2	6.65	Sebarang Ulu I	Bad
30-07-2016	5.06	3	6.54	Sebarang Ulu I	Bad
30-07-2016	4.94	2	5.34	Sebarang Ulu I	Bad
30-07-2016	5.06	2	4.73	Sebarang Ulu I	Bad
30-07-2016	5.06	3	5.70	Sebarang Ulu I	Bad
30-07-2016	4.94	2	6.95	Sebarang Ulu I	Bad
30-07-2016	4.94	3	4.94	Sebarang Ulu I	Bad
30-07-2016	4.94	2	8.89	Sebarang Ulu I	Bad
30-07-2016	4.94	1	8.82	Sebarang Ulu I	Bad
30-07-2016	4.82	1	8.53	Sebarang Ulu I	Bad
30-07-2016	4.70	1	8.22	Sebarang Ulu I	Bad
30-07-2016	4.46	2	8.13	Sebarang Ulu I	Bad
30-07-2016	4.22	2	8.03	Sebarang Ulu I	Bad
30-07-2016	4.10	2	8.48	Sebarang Ulu I	Bad
30-07-2016	4.10	2	4.01	Sebarang Ulu I	Bad
30-07-2016	4.58	4	4	Sebarang Ulu I	Bad
30-07-2016	4.58	1	4.87	Sebarang Ulu I	Bad
30-07-2016	5.06	2	4.03	Sebarang Ulu I	Bad
30-07-2016	5.30	3	3.97	Sebarang Ulu I	Bad
30-07-2016	5.42	2	3.94	Sebarang Ulu I	Bad
30-07-2016	5.30	3	3.94	Sebarang Ulu I	Bad
30-07-2016	5.30	2	3.96	Sebarang Ulu I	Bad
30-07-2016	5.42	2	3.931	Sebarang Ulu I	Bad
30-07-2016	5.42	2	3.946	Sebarang Ulu I	Bad
30-07-2016	5.42	2	3.931	Sebarang Ulu I	Bad
30-07-2016	5.42	2	4.011	Sebarang Ulu I	Bad
29-07-2016	9.26	8	7.57	Kalidoni	Bad
29-07-2016	9.14	8	7.549	Kalidoni	Bad
29-07-2016	9.14	6	6.492	Kalidoni	Bad
29-07-2016	9.4	10	6.072	Kalidoni	Bad
29-07-2016	9.02	8	5.561	Kalidoni	Bad
29-07-2016	9.14	8	5.078	Kalidoni	Bad
29-07-2016	9.14	9	6.086	Kalidoni	Bad
29-07-2016	9.14	8	6.219	Kalidoni	Bad
29-07-2016	9.14	9	6.73	Kalidoni	Bad
29-07-2016	9.02	8	7.18	Kalidoni	Bad
29-07-2016	9.02	8	6.25	Kalidoni	Bad
29-07-2016	9.02	7	5.91	Kalidoni	Bad

Date	pH	Amonisk (%)	Keruh (NTU)	Location	State
29-07-2016	9.02	9	5.82	Kalidoni	Bad
29-07-2016	9.02	10	5.97	Kalidoni	Bad
29-07-2016	9.02	6	6.20	Kalidoni	Bad
29-07-2016	9.02	6	6.41	Kalidoni	Bad
29-07-2016	8.90	6	6.37	Kalidoni	Good
29-07-2016	8.90	7	6.21	Kalidoni	Good
29-07-2016	8.78	7	6.77	Kalidoni	Good
29-07-2016	8.90	8	8.45	Kalidoni	Good
27-07-2016	8.78	6	8.49	Sei Gerong	Good
27-07-2016	8.66	7	8.52	Sei Gerong	Good
27-07-2016	8.30	5	8.16	Sei Gerong	Good
27-07-2016	8.30	6	8.00	Sei Gerong	Good
27-07-2016	7.94	5	7.73	Sei Gerong	Good
27-07-2016	7.94	4	7.69	Sei Gerong	Good
27-07-2016	7.70	4	7.07	Sei Gerong	Good
27-07-2016	7.58	4	7.24	Sei Gerong	Good
27-07-2016	5.90	5	5.92	Sei Gerong	Bad
27-07-2016	6.02	6	6.64	Sei Gerong	Good
27-07-2016	6.02	6	6.77	Sei Gerong	Good
27-07-2016	5.90	6	6.81	Sei Gerong	Good
27-07-2016	6.38	13	6.67	Sei Gerong	Good
27-07-2016	6.38	2	6.42	Sei Gerong	Good
27-07-2016	5.90	4	6.51	Sei Gerong	Bad
27-07-2016	5.90	4	6.47	Sei Gerong	Bad
27-07-2016	5.90	5	6.37	Sei Gerong	Bad
27-07-2016	5.90	6	11.81	Sei Gerong	Bad
27-07-2016	5.78	5	7.38	Sei Gerong	Bad
27-07-2016	5.78	5	7.40	Sei Gerong	Bad

Based on the threshold given in Table 1 and Table 2, from 150 samples, SU 1 is the highest level of pollutions as shown in Figure 3.

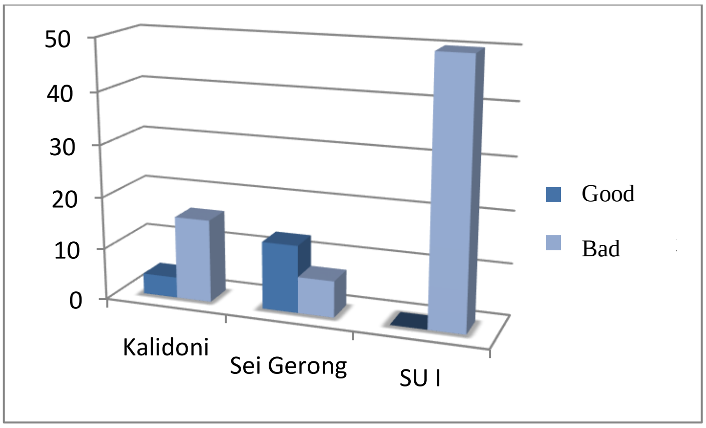

Figure 3. Measurement Results.

6. Conclusion
Based on the research experiments and analysis, it can be conducted the conclusions below:
1. The SVM classification shows good result, but lack justification value. This is because the assessment involves the threshold value only.
2. Seberang Ulu (SU) 1 has significant impact due to the waste pollution.

Acknowledgments
This research was supported by University of Sriwijaya. We thank our colleagues from Computer Science Faculty who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations/conclusions of this paper.

7. References

[1] P. J. Nkosi Nhlanhla, Muzenda Edison, Zvimba John, "The Waste tyre problem in South Africa: An analysis of the REDISA Plan," *Int. Conf. Chem. Environ. Eng.*, no. i, 2013.
[2] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, vol. 47. 2004.
[3] M. A. Ahmadi and M. Pournik, "A Predictive Model of Chemical Flooding for Enhanced Oil Recovery Purposes: Application of Least Square Support Vector Machine," *Petroleum*, vol. 2, no. 2, pp. 177–182, 2015.
[4] M. Malvoni, M. G. De Giorgi, and P. M. Congedo, "Data on Support Vector Machines (SVM) model to forecast photovoltaic power," *Data Br.*, vol. 9, pp. 13–16, 2016.
[5] N. Cristianini and M. H. Hahn, *Introduction to Computational Genomics - A Case Studies Approach*. Cambridge University Press, 2006.