

## Building a Predictive Model to Estimate NOx Emission Pollutant of Backhoe Equipment

Aisyah Larasati<sup>1,\*</sup>, Aal F.P Mahardika<sup>1</sup>, Darin Ramadhanti<sup>1</sup>, Yuh-Wen Chen<sup>2</sup>, Apif Miftahul Hajji<sup>3</sup>, Vertic Eridani Budi Darmawan<sup>1</sup>

<sup>1</sup> Department of Industrial Engineering, Universitas Negeri Malang, Jl. Semarang No.5 Malang - Indonesia

<sup>2</sup> Department of Industrial Engineering and Management, Da-Yeh University, Taiwan

<sup>3</sup> Department of Civil Engineering, Universitas Negeri Malang, Jl. Semarang No.5 Malang – Indonesia

\*Corresponding author : [aisyah.larasati.ft@um.ac.id](mailto:aisyah.larasati.ft@um.ac.id)

**Keywords:** NOx, Support Vector Machine model, SVM

**Abstract.** NOx is one of emission pollutant resulted from Backhoe equipment. This research aims to build a predictive model to estimate NOx pollutant released by backhoe equipment using Support Vector Machine model. Two type of kernel types (radial basis function and linear kernel types) are compared. The study runs the model several time to maximize the accuracy of SVM by finding the optimized parameter, which includes  $C$ ,  $\epsilon$ , and  $\gamma$ . The results show that radial basis function kernel type provide higher accuracy than linear kernel type. In addition, this study also conclude that higher  $C$  and  $\gamma$  parameter results in much lower mean absolute error value. However, it requires much longer calculation time. The SVM predictive model also show that the significant factors to predict NOx emission are MAP, RPM, backhoe type and the intake temperature.

### Introduction

Nowadays automotive technology is moving faster, one of which is in automotive technology for construction vehicle. This phenomena also effects on the productivity of construction. There are many companies make innovation to create construction equipment that gives high productivity. Unfortunately, high productivity of equipment is only assessed through the number of costs incurred without reviewing at the element of a pollutant that will produce [1]. Environmental issues are rarely to be a critical point to design construction equipment with high productivity. Meanwhile, the pollutants produced greatly disrupt the environment and the wider community [2].

NOx is the part of vehicle gas emission according to EMFAC model. The reaction of nitrogen on combustion during high temperature creates NOx. Commonly, NOx is produced by vehicle emission, including construction heavy equipment. NOx and automotive construction equipment have a strong connection. It is undeniable, one of the causes of the construction industry being one of the biggest contributors to gas emissions is from the pollutants produced by the construction automotive equipment. The number of NOx pollutants that exist is directly proportional to the number of automotive equipment available. According to the U.S. Environmental Protection Agency (EPA) data, Heavy Duty Diesel (HDD) construction equipment produce 657,000 tons of NOx emission [3].

There are much automotive equipment related to a construction project, including backhoe. Backhoe is heavy duty diesel (HDD) construction equipment that takes big part on the productivity of construction. Backhoe is used to dig the natural surface of the ground and load it to truck or tractor [4]. Diesel vehicle, including Non-road and On-road type, releases a significant amount of pollutant [5]. Backhoe also becomes one of the contributors to pollutants from construction equipment. There are several components and specification of the backhoe that is related to the number of pollutants produced.

The type of backhoe is one of many aspects related to the amount of pollutant produced. Different types of backhoes have different specifications. By analyzing 6 different types of backhoes, it can be concluded which type of backhoe produces a large number of NO<sub>x</sub> pollutants. Meanwhile, the size of horsepower machine could also influence on backhoe emission. Horsepower and Backhoe type can show the relation between NO<sub>x</sub> emission and characteristic of backhoe machine [6].

In other views, there are RPM, Norm MAP, Engine tier, and temperature who have significantly influence on backhoe emission. RPM (Revolution per Minute) is used to express the speed of rotation of an axis in one minute. In the vehicle, 1 rpm means 1 cycle of crankshaft rotation or crankshaft. RPM has significant correlation with the engine fuel. Meanwhile, the amount of fuel and energy used will certainly affect the NO<sub>x</sub> emissions issued by the backhoe. This sensor MAP (Manifold Absolute Pressure) served to measure the volume of air entering the engine that occurred in the intake manifold. MAP is used to measure incoming air and will be forwarded to regulate incoming fuel and emit NO<sub>x</sub> emissions. the ability and size of MAP on different backhoes will affect the NO<sub>x</sub> pollutants produced [7]. Engine tier is some standard created by U.S. Environmental Protection Agency (EPA) and the California Air Resources Board applies for diesel equipment to restraint amount of emissions that will be produced, This activity already gives reduction NO<sub>x</sub> emission [3].

There are some method that used to approach main goal of this paper. In data mining, Support Vector Machine (SVM) algorithm is one of popular methods for prediction problem. SVM has regression mode to analyse the relation between variables [8]. SVM regression is a useful method to make prediction on an actual data. Trough data prediction, analyser can take a conclusion which variable who the strongest influence on data experiment. On other case, SVM has characteristic called learning capacity and generalization ability [9]. A model must has these two characters for being good model. This is why SVM regression is chosen to be approached method for this study.

This paper aims to build a predictive model to estimate NO<sub>x</sub> pollutant released by Backhoe equipment. This study also determines the best kernel type and its setting parameter that results in the lowest mean absolute error. The lower the mean absolute error, the better the performance of the SVM regression model. By having the results of NO<sub>x</sub> predictions, automotive company and construction industry will be able to improve and give new innovations to the components so that it will produce high productivity with environmentally friendly machines.

## **Research Methodology**

This research used field data measured by the North Carolina State University research team [10]. From the data set, this study used only six types of backhoe with its specification such as age of the machine, the engine tier, engine rated power, also the engine activity including RPM (revolutions per minute), MAP (kPa), and Temperature. The emissions gas produced by backhoe data being used in this research are mass per time (g/s) rates of NO<sub>x</sub>. During data preprocessing phase, the emission gas was transformed into mass per hour (g/hr) rates of NO<sub>x</sub> in order to make it normally distributed. In this research the types of backhoe including its specification and the engine activity are set as independent variables, while the emissions gas is the dependent variables. This research uses a predictive model of Support Vector Machines to estimate NO<sub>x</sub> emission. The performance of the prediction model is assessed on its mean absolute error.

The research begins with data preparation. The input data contains seven independent variables and one dependent variables which is NO<sub>x</sub> emission. The next step is to determine data type of the variables such as engine tier and backhoe are nominal data, while the others variables are scale. The study continues with cleaning the data set, so there is no missing data and outlier data points on all variables. The second step is to build the Support Vector Machine model using data set that have been prepared. Before building the SVM model, the data set are split into two data sets: training and testing. Each of training and testing are comprised of 50 percent respectively. After splitting the data, the next step is to set the parameters. The parameters being used in this research are two kernel types, the RBF (Radian basis function) and Linear Kernel Type, and also the setting parameters for each kernel types such as  $C$ ,  $\epsilon$ , and  $\gamma$ . The last step is to run the model of support vector machine until optimized

condition reached based on the mean absolute error value. Mean absolute error are the reflection of the model accuracy, the lower mean absolute error, the better model is [11]. The model simulation is run for 20 cycles with each different parameter being set. In brief the method of this research is shown in Figure 1.

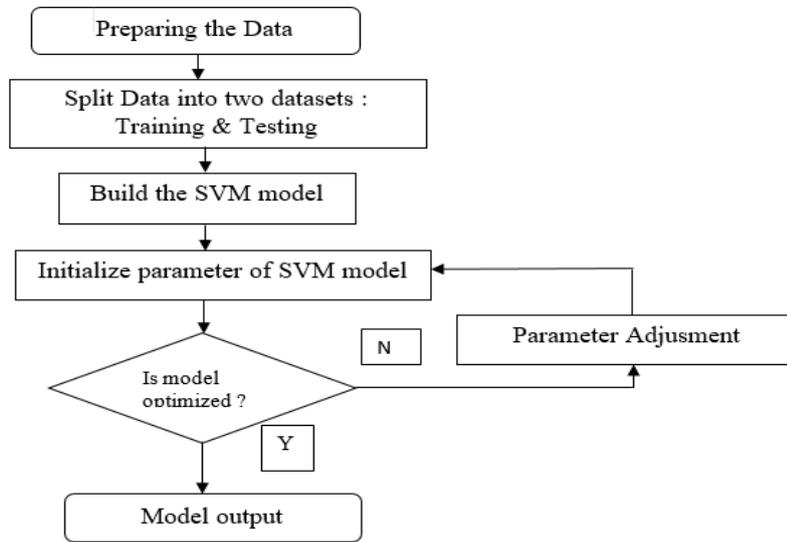


Fig. 1. The Flowchart of Research Method

### Results and Discussion

On SVM model, the characteristics of the model is determined by the kernel function. This study compares the SVM model performance by changing the kernel function between linear kernel and radial basis function (RBF) kernel function. Both of kernel model requires the same parameters, which are C and Epsilon ( $\epsilon$ ). The difference between both kernel types is the existing of gamma parameter in the RBF kernel type, while the linear kernel does not contain gamma parameter. Applying the correct parameter selection leads to good model performance, which is indicated by having low mean absolute error value [12]. The scatter plot of the predicted and actual data of NO<sub>x</sub> studied in this research is shown in Figure 2.

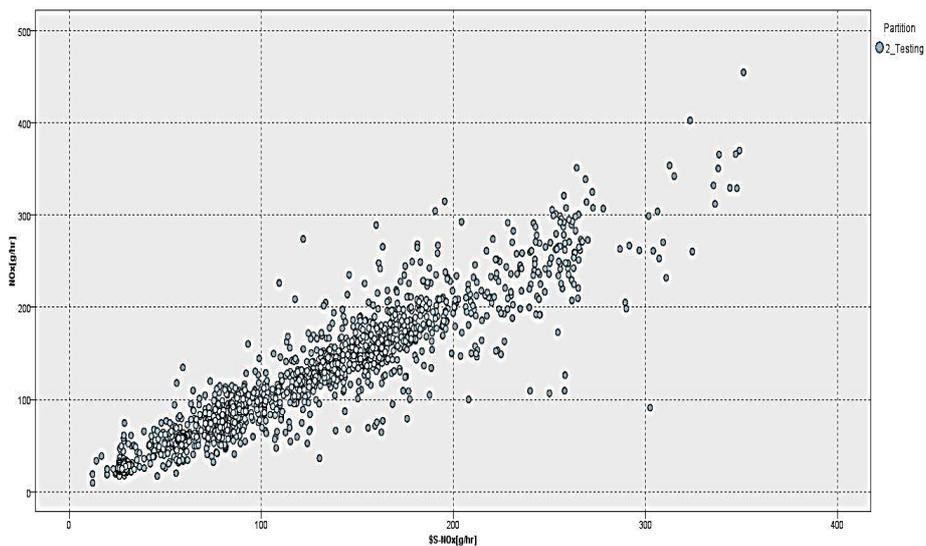


Fig. 2. Scatter Plot Predicted versus Actual NO<sub>x</sub> Emission Data

Based on the scatter plot shown in Figure 2, it can be concluded that bias value of the predicted data using SVM model is relatively low because the predicted data is scattered around actual points. It implies that the bias of the predicted value compared the actual value is relatively low.

The result of the Support Vector Machine predictive model is shown in Table 1. The model were simulated for 20 cycles and each cycles had different parameter set up. Table 1 show seven cycles of the SVM model using RBF kernel type and five cycles of the SVM model using Linear kernel type. The other eight cycles do not listed in Table 1 because there is no difference in the mean absolute error value obtained from the model.

Table 1. Mean Absolute Error Value

| Parameter<br>Radian Basis Function                      | Mean Absolute<br>Error | Parameter Linear           | Mean Absolute Error |
|---|------------------------|----------------------------|---------------------|
| $C = 10 ; \epsilon = 0.1 ; \text{RBF } \gamma = 0.1$    | 22.35                  | $C = 10 ; \epsilon = 0.1$  | 26.251              |
| $C = 40 ; \epsilon = 0.05 ; \text{RBF } \gamma = 0.5$   | 18.29                  | $C = 10 ; \epsilon = 0.05$ | 26.251              |
| $C = 70 ; \epsilon = 0.05 ; \text{RBF } \gamma = 0.9$   | 17.19                  | $C = 40 ; \epsilon = 0.05$ | 26.246              |
| $C = 100 ; \epsilon = 0.1 ; \text{RBF } \gamma = 1.5$   | 16.35                  | $C = 40 ; \epsilon = 0.1$  | 26.247              |
| $C = 800 ; \epsilon = 0.1 ; \text{RBF } \gamma = 8.0$   | 14.49                  | $C = 40 ; \epsilon = 0.1$  | 26.247              |
| $C = 1000 ; \epsilon = 0.1 ; \text{RBF } \gamma = 10.0$ | 14.35                  | -                          | -                   |
| $C = 1500 ; \epsilon = 0.1 ; \text{RBF } \gamma = 15.0$ | 14.15                  | -                          | -                   |

Results in Table 1 show that SVM models built by using linear kernel type do no result in significant differences on its mean absolute error for each simulation cycle. The minimum mean absolute error value for linear kernel type function is 26.246, which is obtained when the parameter  $C = 40$  and from  $C = 10$  and  $\epsilon = 0.05$ . The advantage of linear kernel function compared to RBF kernel type function is the kernel function is simpler since it requires only one parameter to be set up, which is regularization  $C$  [13]. Linear kernel is more effective and faster to be optimized than RBF kernel type function. However, linear kernel type function does not always deliver optimum condition, thus this study continues building SVM model using the RBF kernel type function.

As have been shown in Table 1, this study build SVM model with RBF kernel in seven cycles since when the  $C$  parameter increases, the mean absolute error decreases. In this study, the SVM model with RBF kernel function have been simulated for 20 cycles. The value of mean absolute error optimal on 14.15 with given parameter  $C = 1500$ , and  $\text{RBF } \gamma = 10$  and the epsilon is kept constant. The first cycle with simple parameter given  $C = 10$  and  $\text{RBF } \gamma = 0.1$  results in mean absolute error value at 22.35. The  $C$  parameter has a function to adjust the confidence interval meanwhile  $\gamma$  is changing the mapping function and complexity in certain datasets (training and testing) [14]. Generally  $C$  is a parameter which is balancing the misclassification in the training data examples to the simplicity of the decision function. Manually, finding the optimal value for the parameters are using grid search on the set up parameters [15]. But, it won't be able to change the parameters whatever it likes. Because of the aim of this research is applying SVM and finding the best accuracy by simulating it many times.

Determining the parameters, such as  $C$  and  $\epsilon$ , are clearly given to users based on literature or user experience [16]. It is not only based on the user experience on changing the parameters value, but also the number of the training datasets, especially on the RBF kernel type. In this research the ratio for the training and testing datasets are 50:50. As have been explained before,  $C$  and  $\epsilon$  is one of the parameters given in RBF kernel type.  $C$  can determines the balancing of the flatness of the model and to see how far the deviations are tolerated. In the case  $C$  value is too large, the prediction risk will be minimized. It can be seen in Table 1, on cycles 7  $C$  is raised into 1500 from 100 in cycles 4. It indicates that the mean absolute error value is already close to the optimal value. On the other hand,  $C$  values above 1500 does not give any higher mean absolute error value. It means the model only give minor effect on the prediction risk [17]. Based on the results shown in Table 1,  $\epsilon$  values does not have any

significant effect, either on  $\epsilon$  values 0.05 or 0.1. Epsilon or ( $\epsilon$ ) small values correspond to higher percentage of support vectors, while C has unnecessary effect on support vectors percentage [17].

In addition, on linear kernel type, C values does not go higher than 40. It indicates that the SVM model using linear kernel type already balance and close to the optimal value. This condition gives higher prediction risk for the support vectors value. Besides the two parameters, there is gamma ( $\gamma$ ) value, which is only shown on RBF kernel but not in linear kernel. Gamma parameters defines the influence of training datasets and how far the models reach it. This implies the inverse of the influence of training datasets selected by the support vectors. By any chance, if this parameter is set at the value that is too large, the model can be over fitted because no amount of C value will prevent it. In Table 1, gamma value does not go higher than 15, because it may cause a separation hyper plane with a small sized margin [18]. This means by setting up the C and  $\gamma$  parameter higher, the model tends to have a lower mean absolute error value. The importance of the predictor variables in the optimized model using radial kernel function type is shown Figure 3.

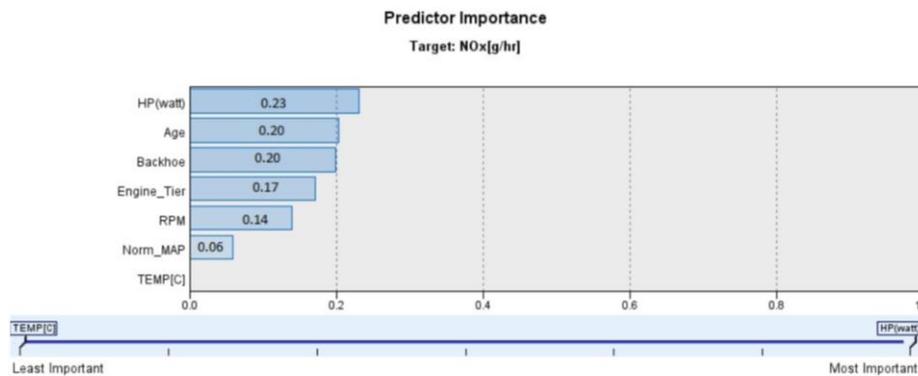


Fig. 3. The Predictor Importance of the Optimized SVM Model

Figure 3 indicates that at the optimum condition (lowest mean absolute error), engine rated power of backhoe becomes the most important predictor variable and the intake temperature becomes the least important. The backhoe age and backhoe type have the same important level. These results explains that the higher the rated power of the backhoe equipment the higher the NOx emission. However, different environment temperature does not really influence the NOx emission. The construction industry should consider the age of its equipment and choose backhoe types carefully in order to minimize NOx emission.

Engine rated power or horsepower is machine ability to carry a load within a certain period of time or the amount of energy that can be released or produced by a machine in a certain time. In automotive technology, horsepower used to accelerate vehicle speed at a steady pace when they going faster every minute. Horsepower has a significant relationship with the speed of the vehicle. Horsepower indicates the machine power and be used to classify type of the machine specification.

In this paper, there are five types of backhoe. Backhoe type is divided based on the machine specification, one of these is the variability of the horsepower. Type 1, 2 and 3 have 65621.6 hp specification with various engine tier and age. Type 4 has 72332.9 hp, and type 7 has 73824.3 hp. The variability of horsepower brings a big impact on the amount of NOx emission released. This is proofed on the result of predictor importance as shown in Figure 3. In this case, horsepower gives 0.23 or 23% importance level, which have the biggest predictor importance than the other specification and indicators.

In reality, Horsepower indeed has significant relation on vehicle emission release. Horsepower describes the strength of the machine. The bigger the power of the machine, the bigger the amount of fuel consumption. The vehicle fuel consumption creates gas emission in their combustion process. Therefore, some researcher makes innovation on horsepower for getting lower maximum output, so the vehicle ran release minimal gas emission [19]. This innovation shows that horsepower or engine rated power give a significant impact on gas emission, including NOx.

## Conclusion

The SVM predictive model show that at the optimum condition (lowest mean absolute error), engine rated power of backhoe becomes the most important predictor variable and the intake temperature becomes the least

important. The backhoe age and backhoe type have the same important level. These results explain that the higher the rated power of the backhoe equipment the higher the NO<sub>x</sub> emission. However, different environment temperature does not really influence the NO<sub>x</sub> emission. The construction industry should consider the age of its equipment and choose backhoe types carefully in order to minimize NO<sub>x</sub> emission.

Moreover, this study also reveals that radial basis function kernel type provide higher accuracy than linear kernel type. RBF kernel had given lower mean absolute error than linear kernel. Higher value of parameter  $C$  and  $\gamma$  results in lower mean absolute error on testing datasets. In contrast, higher value of  $C$  and  $\gamma$  requires much longer calculation time. The calculation time increases exponentially following the increase in the value of  $C$  and  $\gamma$ .

## Acknowledgement

Acknowledgments to Universitas Negeri Malang (UM) which has funded this research through PNBPNP - UM Research Grant 2019.

## References

- [1] Hajji AM, Larasati A, Lewis MP. Combining Off-the-Job Productivity Regression Model with EPA's NONROAD Model in Estimating CO<sub>2</sub> Emissions from Bulldozer. *Civ Eng Dimens* 2017;19:73–8. doi:10.9744/ced.19.2.73-78.
- [2] Hajji AM, Lewis MP. How to estimate green house gas (GHG) emissions from an excavator by using CAT's performance chart. *AIP Conf Proc* 2017;1887. doi:10.1063/1.5003530.
- [3] Lewis P, Fitriani H, Arocho I. Engine variable impact analysis of fuel use and emissions for heavy-duty diesel maintenance equipment. *Transp Res Rec* 2015;2482:8–15.
- [4] Vaniya HK. Analysis of Backhoe Loader Chassis for Weight & Cost Reduction using FEA - A Review Paper 2015;1:1–8.
- [5] Frey HC, Rasdorf W, Kim K, Pang S, Lewis P. Comparison of Real-World Emissions of B20 Biodiesel versus Petroleum Diesel for Selected Nonroad Vehicles and Engine Tiers. *Transp Res Rec J Transp Res Board* 2008;2058:33–42. doi:10.3141/2058-05.
- [6] Kacem M, Zaghdoudi K, Morales-Rubio A, de la Guardia M. Preliminary results on the influence of car characteristics on their gases emissions using gas sensors. *Microchem J* 2018;139:69–73. doi:10.1016/j.microc.2018.02.022.
- [7] Park Y, Min K, Chung J. Control of the air system of a diesel engine using the intake oxygen concentration and the manifold absolute pressure with nitrogen oxide feedback 2016. doi:10.1177/0954407015584130.
- [8] Kwiatkowski A, Chludzinski T, Smulko J. Metrology and measurement systems. *Metrol Meas Syst* 2018;25:551–60. doi:10.24425/123892.
- [9] Smits GF, Jordaan EM. Improved SVM regression using mixtures of kernels 2003:2785–90. doi:10.1109/ijcnn.2002.1007589.
- [10] Lewis P, Rasdorf W. Fuel Use and Pollutant Emissions Taxonomy for Heavy Duty Diesel Construction Equipment. *J Manag Eng* 2017;33.
- [11] Xia L, Pan H. Inferential Estimation of Polypropylene Melt Index Using Stacked Neural Networks Based on Absolute Error Criteria 2010:216–8.
- [12] Chang Y, Lin C. Training and Testing Low-degree Polynomial Data Mappings via Linear SVM 2010;11:1471–90.
- [13] Spratt H, Ju H, Brasier AR. A structured approach to predictive modeling of a two-class problem

- using multidimensional data sets. *Methods* 2013;61:73–85. doi:10.1016/j.ymeth.2013.01.002.
- [14] Han S, Qubo C, Meng H. Parameter selection in SVM with RBF kernel function. *World Autom. Congr.* 2012, IEEE; 2012, p. 1–4.
- [15] Lauzana I, Platerrier B. *Advanced Machine Learning Practical 3 Solution : Classification Non-linear Classification Techniques* 2017:1–31.
- [16] Cortes C, Vapnik V. *Support-Vector Networks* 1995;297:273–97.
- [17] Cherkassky V, Ma Y. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks* 2004;17:113–26.
- [18] Ding I, Shi J. Hybridized estimations of support vector machine free parameters  $C$  and  $\gamma$  using a fuzzy learning strategy for microphone array-based speaker recognition in a Kinect sensor-deployed environment 2017. doi:10.1007/s11042-017-4499-y.
- [19] Kawamoto N, Naiki K, Kawai T, Shikida T, Tomatsuri M. Development of new 1.8-liter engine for hybrid vehicles. *SAE Technical Paper*; 2009.