

Carbon dioxide emission prediction using support vector machine

Chairul Saleh¹, Nur Rachman Dzakiyullah², Jonathan Bayu Nugroho³

¹Department of Industrial Engineering, Faculty of Industrial Technology, Universitas Islam Indonesia, Yogyakarta, Indonesia

²Faculty of Information Technology and Communication Universiti Teknikal Malaysia Melaka (UTEM) Melaka, Malaysia

³Department of Industrial Engineering, Faculty of Industrial Technology, Universitas Islam Indonesia, Yogyakarta, Indonesia

Abstract. In this paper, the SVM model was proposed for predict expenditure of carbon (CO₂) emission. The energy consumption such as electrical energy and burning coal is input variable that affect directly increasing of CO₂ emissions were conducted to built the model. Our objective is to monitor the CO₂ emission based on the electrical energy and burning coal used from the production process. The data electrical energy and burning coal used were obtained from Alcohol Industry in order to training and testing the models. It divided by cross-validation technique into 90% of training data and 10% of testing data. To find the optimal parameters of SVM model was used the trial and error approach on the experiment by adjusting C parameters and Epsilon. The result shows that the SVM model has an optimal parameter on C parameters 0.1 and 0 Epsilon. To measure the error of the model by using Root Mean Square Error (RMSE) with error value as 0.004. The smallest error of the model represents more accurately prediction. As a practice, this paper was contributing for an executive manager in making the effective decision for the business operation were monitoring expenditure of CO₂ emission.

1. Introduction

The global warming issues now become a universal problem for all nations. The Intergovernmental Panel on Climate Change (IPCC) reported that scientists were more than 95% certain that most of global warming is caused by increasing concentrations of greenhouse gasses and other human (anthropogenic) activities [1]. That balance between earth and atmosphere affected by an increase in acid gasses charcoal or carbon dioxide (CO₂), methane (CH₄), nitrous oxide (N₂O), hydrofluorocarbons (HFC) and perfluorocarbons (PFC) more commonly known by greenhouse gasses. In particular, CO₂ is a major cause of Global Warming [2]. About eight billion tons per year of carbon in the form of CO₂ emitted globally through burning fossil fuels for transport and for the production of heat and electricity around the world [3]. Emission of carbon dioxide is the rest results of combustion of water (H₂O) and carbon monoxide gas (CO) or also called as carbon dioxide (CO₂) that is a greenhouse gas. There's a concept as references in the measurement of CO₂ emission, namely carbon footprint [4].

¹ Corresponding Author: Chairul@uii.ac.id (Chairul Saleh)

² Nurrachmandzakiyullah@gmail.com

³ jbayunugroho@gmail.com



Carbon footprint has become a widely used term and concept in the public debate on responsibility and abatement action against the threat of global climate change. It had a tremendous increase in public appearance over the last few months and years and is now a buzzword widely used across the media, the government and in the business world [5]. It is a measure of total amount of carbon dioxide released into the atmosphere in the given time frame that is directly or indirectly caused by an activity to provide service or product [6]. A carbon footprint is a measure of the amount of carbon dioxide emitted through the combustion of fossil fuels. In the case of a business organization, it is the amount of CO₂ emitted either directly or indirectly as a result of its everyday operations. It also might reflect the fossil energy represented in a product or commodity reaching market [7]. The carbon footprint of U.S. households is about 5 times greater than the global average, which is approximately 10 tons CO₂e per household per year. For most U.S. households, the single most important action to reduce their carbon footprint is driving less or switching to a more efficient vehicle [8].

During the last decade, the growth of energy consumption and CO₂ emissions be more important than industry growth trends, it means that the reduction of energy consumption and CO₂ emission or slightly does not cause a decrease in growth industries. The industrial sector is the main contributor of the total CO₂ emissions in the world [9]. Considering the growth of manufacturing sector, expected there is serious concern about the emission of carbon dioxide [10]. The use of energy is the greatest source of emissions. This represents about 65 percent of all gas emissions, and the level of emission is expected to rise if no accurate steps implemented. Coal and Electricity have been a preferred form for energy consumption and has consistently registered a higher growth rate than other forms of energy. Increased consumption of electrical power is more intimately bound up with increased emission levels of CO₂ [11]. Recently, many efforts have been put on the complex system of economy-environment and relevant CO₂ emission reduction issues [12]. In an effort to minimizing CO₂, there is a need of new manufacturing process, better known as green manufacturing which is suitable a sustainable development strategic [13]. Green manufacture is an economically-driven, system-wide, and integration approach to the reduction and elimination of all waste streams associated with the design, manufacture, use and disposal products and materials [14]. The standards to reach green manufacture including zero potential safety problems, zero health threats on the operators and product users, zero environmental pollution, waste recycling, and waste disposal during the production process as much as possible [15].

2. Literature Review

Research on CO₂ emissions is highly renewable, qualitative and quantitative researches that have been done until now is still in the process of the discussion, it makes the theory and approaching method in order to calculate the emitter of CO₂ emissions has not become one unit [16]. Mitigation of Carbon Dioxide emission is the challenge of the future in order to stabilize global warming [17]. CO₂ prediction using computational intelligence approach that has been done by [18]. An adaptive neuro-fuzzy interference system (ANFIS) and multi-layer perceptron artificial neural network (MLP-ANN) have been developed to estimate CO₂. The proposed model of ANFIS and MLP-ANN demonstrates that both methods can solve CO₂ prediction problem. There's another method for prediction namely, support vector machine (SVM). Recently, several applications of SVM can be found both for classification and for regression problems [19]. SVM is more accurate than semi empirical equations for predict solubility of different solutes in supercritical carbon dioxide [20]. On the other hand, SVM was implements the structural risk minimization principle [21]. SVM has successfully solved a prediction of CO₂ exchange rate [22]. In their study, different tests were performed along the North Atlantic oceanic region with data obtained during 2009 and the proposed model of SVM demonstrates that SVM can solve CO₂ prediction problem. SVM is powerful machine learning tool that can be used for time-series prediction [26].

SVM is a novel learning machine first developed by Vapnik in 1995 [23–25]. SVM is a learning system that uses a hypothetical space in the form of linear functions in a high dimension feature space, trained with the learning algorithm based on the theory of optimization by implementing learning bias. SVM concept uses the concept of ϵ -incentive function loss. SVM can be generalized to approach non-linear function known as SVR. SVM concept can be defined as using single hyperplane in many

dimensional spaces that eventually those partitions can be resolved in a non-linear. SVM considers the two-class classification, the classes being $y_i = +1, -1$ respectively. To separate the training pattern in both classes, SVM will find the maximum hyperplane. Linear equation of hyperplane where the hyperplane normal to w and b is bias.

$$w^T \varphi(x) + b = 0 \quad (1)$$

If the training data are linearly separable then a pair of (w, b) will be existed as

$$w^T \varphi(x) + b = -1 \quad (2)$$

$$w^T \varphi(x) + b = +1 \quad (3)$$

If the classes is not linearly separable then a non-negative slack vector variable $\xi = (\xi_1, \dots, \xi_m)$ will be existed as

$$w^T x_i + b + \xi_i = -1 \quad (4)$$

$$w^T x_i + b + \xi_i = +1 \quad (5)$$

Based on many researches mentioned above, it is proved that there are so many good methods to predict the carbon dioxide emission. In this research will be carried out the prediction of carbon dioxide in the manufacturing industry. Machine learning of support vector machine approach will be used as a prediction method.

3. Proposed SVM model for carbon dioxide emission prediction

This paper will investigate CO₂ emissions using prediction SVM model, which considering variable energy consumption that have an impact on the emergence of CO₂ emission. In this study, we considering electrical energy and burning coal as energy consumption that affect directly increasing of CO₂ emissions [11]. The data of energy consumption were collected from the alcohol industry in Yogyakarta, Indonesia. The data of energy consumption is retrieved from historical data that used in production process, including electrical energy during the production process in kilowatt-hour (kWh), burning coal in kilogram (Kg). However, the data electrical energy and burning coal consumption must be converted into the values of CO₂ emission (KgCO₂). The value of the emission factor for electricity refers to the emission factors that have been set by Perusahaan Listrik Negara (PLN). PLN is a state's company that deal with all aspects of electricity in Indonesia. The value of the emission factor of the electricity that is equal to 0.725 kg/kWh. Conversion of electrical energy is obtained by multiplying the amount of energy used and the calorific value or emission factor as shown in Equation 6.

$$E_{CO_2}^E = \sum_{t=1}^{t=} A_{(E)t} \times F_{(E)t} \quad (6)$$

Where,

$E_{CO_2}^E$ = Total CO₂ emission from electrical energy consumption (kg.CO₂)

$A_{(E)t}$ = Electrical energy used (kWh)

$F_{(E)t}$ = Electrical emission factor (kg/kWh)

Then, the amount of CO₂ emissions combustion of fossil fuels (coal) results depends on many and the type of fuel burned. The amount of fuel represented as activity data while fuel types are represented by the emission factors. The general equation used to estimate CO₂ emissions from coal are as follows:

$$E_{CO_2}^F = C \times EF \quad (7)$$

$E_{CO_2}^F$ = CO₂ emission from combusted fuel (Kg.CO₂)

C = Coal Energy consumption (TJ)

EF = Emission factor (Kg/TJ)

Where, C must be converted first as

$$\begin{aligned} C &= Q \times P_v \\ Q &= \text{Quantity of Coal energy used (Kg)} \\ P_v &= \text{Calorific Value (TJ/Kg)} \end{aligned}$$

Then, for the CO₂ emission data are obtained the total emitted CO₂ from electrical energy consumption and coal. In order to designing prediction model of SVM, training and testing data were needed. In this research, electrical energy consumption and coal data were selected as dataset input and the CO₂ emission as output. The relationship between the input and output variables can be defined as follow:

$$\left[x_i \begin{bmatrix} \text{Electrical Energy} \\ \text{Burning Coal} \end{bmatrix} \right] y_i = [\text{Carbon Dioxide Emission}] \quad (3)$$

Furthermore, to make it clearly, there are several step in order to design model prediction of SVM as shown in figure 1 as follow:

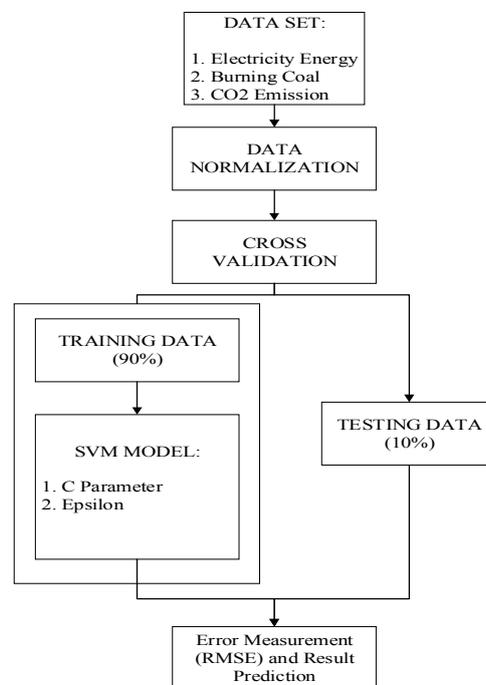


Figure 1. Design Model of SVM

However, in order to design prediction SVM model pre-processing data is needed by data mining technique that involves transforming raw data into the understandable and proper format. The pre-processing data is normalization that needed to improve the generalization performance with SVM [27]. Then, to validate the model, cross-validation is used by dividing the dataset into training data (90%) and testing data (10%). The model proposed in this research is used to support vector machine algorithm to investigate the learning method for the CO₂ emission prediction. A learning process in SVM is used to Support Vector Regression (SVR). SVR is a part of Support Vector Machine and is specialized in obtaining regression models by means of a change in the dimensionality of the data. SVR concept is based on risk minimization, i.e. to estimate a function by minimizing the upper limit of generalization error, so that the SVR is able to overcome the overfitting. To find the optimal parameters of the SVM

model in the training process, the parameters such as C parameter, Epsilon must be optimal. To measure the optimal parameter of prediction SVM model will be used root mean square error (RMSE) [28]. The obtained result of RMSE shows the accuracy of the prediction. Then, the result prediction will be analyzed.

4. Experiment result

In this work, the application software is Rapid Miner 5.3 was conducted in the experiment for predict CO₂ emission. In order to test problems were undertaken using a computer specification: AMD A8-5550M 2,1 GHz, RAM 2 GB and Windows 10 as the operating system. Our objective is to monitor CO₂ emissions from incurred energy consumption by looking for an accurate prediction which has the lowest error. In this experiment, as mention before, cross-validation was used to split the data into a training set or equal parts: 90% and 10 % for testing set. Then, Error measurement was performed on 90/10 splits and repeated for all 10 probable splits. In order to get the optimal parameters of SVM model, the SVR learning with a trial and error approach for the training process was performed. The objective of SVM model is finding the best parameters on SVM model such as C parameters, Epsilon and the type kernel function used “dot” function. We set the maximum iteration 10.000 and during the training, C parameters were determining -0.1 until 1 and 2 until 50. The parameter of C selected based on smallest RMSE value. Then, after C parameter is obtained, Epsilon was determined with range 0.1 until 1. The best epsilon parameters were selected also based on smallest RMSE value. The C parameters are used to minimize error during the training process and can avoid overfitting and underfitting if the parameter was properly selected. Overfitting arises whenever the training algorithm capturing noise from the data, while underfitting arises if the training algorithm cannot accurately recognize the underlying trend of the data, therefore the model shows low bias but high variance. Moreover, the Epsilon is used to fit the training data. Both parameters can be selected by the user [29]. In Table 1 shows the optimal parameters of SVM model for predict CO₂ as follow:

Table 1. Optimal Parameter on CO₂ emission prediction model

C	Epsilon	RMSE
0.1	0	0.004

Based on the experiments result above, show that the optimal parameter of SVM is C is 0.1 and epsilon are 0. The performance of the SVM model has an error value (RMSE) of 0.004. By using a parameter in Table 1, we compare the actual data and the result of prediction CO₂ emission as follow.

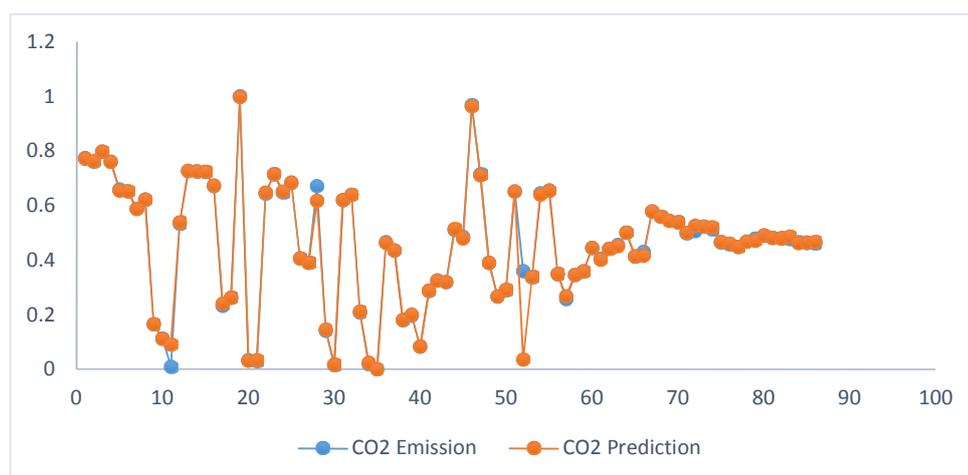


Figure 2. Results of actual CO₂ emission VS CO₂ emission prediction

In the last step, the output of the SVM model is analyzed. To ensure the prediction result are a normal distribution, the confidence level of training model SVM need to be done by using the statistical technique. Several statistic tests can be used to test the normal distribution such as Shapiro–Wilk test, the Lilliefors test, D’Agostino–Pearson’s L2 test, the Jarque–Bera test, and the Anderson–Darling test [30]. However, normal distribution test Shapiro–Wilk was conducted because the greatest applicable statistic method that suitable for all type of distribution and sample in order to test the model prediction. If the prediction value was normally distributed, it shows that the prediction model has higher accuracy and precision. We set the significance level (α) at 0.05; if $p < 0.05$, then the prediction value are not normally distributed. The result of the Shapiro–Wilk test is shown in Table 2.

Table 2. Normality Test

Shapiro-Wilk		
Statistic	df	Sig.
0.976	86	0.104

As can be seen from Table 2, prediction data have the significance value is 0.104 respectively. Thus, the predicted values are normally distributed. Moreover, histogram analysis for prediction data as shown in Figure 3. it indicates that the prediction data have higher accuracy.

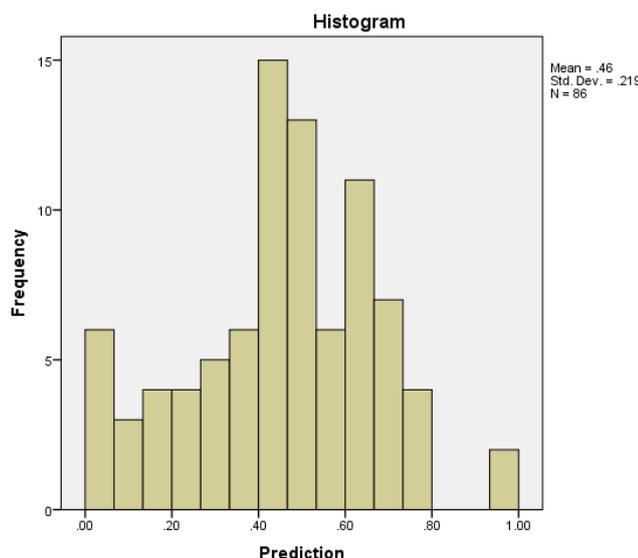


Figure 3. Histogram analysis of Normal distribution

The statistical analysis generally used to know how the statistical probability distribution of the data which can help to make the right decision. From Figure 3, the prediction results approximate a normal distribution were the histogram is centered over the true value. It proves that the SVM model can solve nonlinear data and higher accuracy of prediction.

5. Conclusions and Further Work

Support Vector Machine model was applied in this paper to predict CO2 emissions from energy consumption. The model is used to monitor electrical energy and burning coal which affect the amount of CO2 emitted. Trial and error approach was applied in order obtain a better prediction model with a lower error. The results obtained show that the lower error (RMSE) value was 0.004 with optimal parameters for the SVM model of 0.1 for the C parameter and 0 for Epsilon. Prediction with high accuracy can give information concerning about CO2 emissions. Furthermore, the main objective in this

work is to achieve the lower RMSE when designing the model prediction. It can be concluded that when the high accuracy of the prediction model, then the lower RMSE value must be obtained. By monitoring energy consumption, it can help the manager to develop policies or taking a decision in order to reduce the negative impact on the environment during the production process. For further research, the parameter of SVM model can be automatically selected by integrating optimization technique such as genetic algorithm or particle swarm optimization.

Acknowledgement

The authors would like to thank the anonymous reviewers for their valuable comments. This research is supported by Directorate of Research and Community Service and Board Academic Development, Universitas Islam Indonesia.

References

- [1] IPCC 2014 Climate Change 2014 Synthesis Report: Headline statements from the Summary for Policymakers
- [2] IPCC 2007 Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change
- [3] Dayaratne S P and Gunawardana K D 2014 Carbon footprint reduction: a critical study of rubber production in small and medium scale enterprises in Sri Lanka *J. Clean. Prod.*
- [4] Wiedmann T, Minx J, Barrett J and Wackernagel M 2006 Allocating ecological footprints to final consumption categories with input-output analysis *Ecol. Econ.* **56** 28–48
- [5] Wiedmann T and Minx J 2007 A Definition of ‘ Carbon Footprint *Science (80-.)*. **1** 1–11
- [6] Andrić I, Jamali-Zghal N, Santarelli M, Lacarrière B and Le Corre O 2015 Environmental performance assessment of retrofitting existing coal fired power plants to co-firing with biomass: carbon footprint and emergy approach *J. Clean. Prod.* **103** 13–27
- [7] Grubb T and Ellis C 2007 Meeting the carbon challenge: The role of commercial real estate owners *Users Manag. Chicago*
- [8] Jones C M and Kammen D M 2011 Quantifying carbon footprint reduction opportunities for US households and communities *Environ. Sci. Technol.* **45** 4088–95
- [9] Oh D-Y, Noguchi T, Kitagaki R and Park W-J 2014 CO₂ emission reduction by reuse of building material waste in the Japanese cement industry *Renew. Sustain. Energy Rev.* **38** 796–810
- [10] Lin B, Moubarak M and Ouyang X 2014 Carbon dioxide emissions and growth of the manufacturing sector: Evidence for China *Energy* **76** 830–7
- [11] Raghuvanshi S P, Chandra A and Raghav A K 2006 Carbon dioxide emissions from coal based power generation in India *Energy Convers. Manag.* **47** 427–41
- [12] Chang N 2015 Changing industrial structure to reduce carbon dioxide emissions: a Chinese application *J. Clean. Prod.* **103** 40–8
- [13] Paul I D, Bhole G P and Chaudhari J R 2014 A Review on Green Manufacturing: It’s Important, Methodology and its Application *Procedia Mater. Sci.* **6** 1644–9
- [14] Sezen B and Çankaya S Y 2013 Effects of green manufacturing and eco-innovation on sustainability performance *Procedia-Social Behav. Sci.* **99** 154–63
- [15] Gao Y, Li J and Song Y 2009 Performance evaluation of green supply chain management based on membership conversion algorithm *Computing, Communication, Control, and Management, 2009. CCCM 2009. ISECS International Colloquium on vol 3* pp 237–40
- [16] Dong Y, Xia B and Chen W 2014 Carbon footprint of urban areas: An analysis based on emission sources account model *Environ. Sci. Policy* **44** 181–9
- [17] Kainuma M, Matsuoka Y and Morita T 2000 The AIM/end-use model and its application to forecast Japanese carbon dioxide emissions *Eur. J. Oper. Res.* **122** 416–25

- [18] Baghban A, Ahmadi M A and Hashemi Shahraki B 2015 Prediction carbon dioxide solubility in presence of various ionic liquids using computational intelligence approaches *J. Supercrit. Fluids* **98** 50–64
- [19] Pinto T, Sousa T M, Praça I, Vale Z and Morais H 2015 Support Vector Machines for decision support in electricity markets' strategic bidding *Neurocomputing*
- [20] Mehdizadeh B and Movagharnejad K 2011 A comparison between neural network method and semi empirical equations to predict the solubility of different compounds in supercritical carbon dioxide *Fluid Phase Equilib.* **303** 40–4
- [21] Protić M, Shamshirband S, Petković D, Abbasi A, Mat Kiah M L, Unar J A, Živković L and Raos M 2015 Forecasting of consumers heat load in district heating systems using the support vector machine with a discrete wavelet transform algorithm *Energy* **87** 343–51
- [22] De Paz J F, Pérez B, González A, Corchado E and Corchado J M 2010 A Support Vector Regression Approach to Predict Carbon Dioxide Exchange *Distributed Computing and Artificial Intelligence* (Springer) pp 157–64
- [23] Kim K 2003 Financial time series forecasting using support vector machines *Neurocomputing* **55** 307–19
- [24] Sajjan K S, Kumar V and Tyagi B 2015 Genetic algorithm based support vector machine for on-line voltage stability monitoring *Int. J. Electr. Power Energy Syst.* **73** 200–8
- [25] Hashemkhani M, Soleimani R, Fazeli H, Lee M, Bahadori A and Tavalaeian M 2015 Prediction of the binary surface tension of mixtures containing ionic liquids using Support Vector Machine algorithms *J. Mol. Liq.* **211** 534–52
- [26] Wang W, Men C and Lu W 2008 Online prediction model based on support vector machine *Neurocomputing* **71** 550–8
- [27] López F J M, Puertas S M and Arriaza J A T 2014 Training of support vector machine with the use of multivariate normalization *Appl. Soft Comput.* **24** 1105–11
- [28] Zhang Y and Zhang P 2015 Machine training and parameter settings with social emotional optimization algorithm for support vector machine *Pattern Recognit. Lett.* **54** 36–42
- [29] Cherkassky V and Mulier F M 2007 *Learning from data. Concepts, theory and methods*
- [30] Razali N and Wah Y 2011 Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests *J. Stat. Model. Anal.* **2** 21–33