International Journal of Science and Engineering(IJSE)

Home page: http://ejournal.undip.ac.id/index.php/ijse

# Prediction of Weekly Rainfall in Semarang City Use Support Vector Regression (SVR) with Quadratic Loss Function

Alan Prahutama[1)] and Hasbi Yasin[2)]

[1,2)]Department of Statistics, Faculty of Science and Mathematics, Diponegoro University
Jl. Prof. Soedharto SH, Tembalang, Semarang 50275

Email: alan.prahutama@undip.ac.id

***Abstract*** *- Semarang city is one of the busiest city in Indonesia. Doe to its role as the capital city of Central Java, Semarang is known as having a relativity high rate economic activities. The geographic of Semarang city bordered by the Java sea, thus whenever the rainfall is high, there could be flood at certain area. Therefore, prediction of rainfall is very important. Support vector machine (SVM) is one of the most popular methods in nonlinear approach. One of the branches of this method for prediction is support vector regression (SVR). SVR can be approached by quadratic loss function. The study is focus on Semarang rainfall prediction during 2009 to 2013 using several kernel function. Kernel Function can provide optimal weight Some of kernel functions are linear, polynomial, and Radial Basis Function (RBF). Using this method, the study provide 71.61% R-square in the training data, for C parameter 2 with polynomial (p=2), and 71.46% R-square for the testing data.*

Keywords***—Complete feed, sheep, digestibility, palm plantation, palm oil by-product***

Submission: February 1, 2015 Corrected : March 8, 2015 Accepted: June 1, 2015

**Doi**: 10.12777/ijse.9.1.13-16

*[**How to cite this article**:* Prahutama, A. and Yasin, H. *(2015). Prediction* of Weekly Rainfall in Semarang City Use Support Vector Regression (SVR) with Quadratic Loss Function*, International Journal of Science and Engineering, 9(1),13-16.* **Doi**: 10.12777/ijse.9.1.13-16*]*

## I. INTRODUCTION

As a tropical country, Indonesia has high intensity rainfall. Rainfall is part of the climate that can't be controlled. It is measured by water's volume in the earth on certain days, weekly, monthly or yearly. Whereas, rain is natural occurrence which water fall into surface of the earth, as a results of consideration from water particles in the clouds (Nawawi, 2001). According BMKG, there are five criteria of rainfall among others very heavy (more than 100mm), heavy (50-100mm), medium (20-50mm), light (5-20mm) and very light (less than 5mm). The high intensity rainfall can be cause natural disaster such as floods and landslides. It can't be controlled, but we can predict it. Therefore, unanticipated floods represent the most destructive natural hazard to threaten human life and properties. Accurate predict of rainfall information is the best approach to avoid life losing and economic loses.

Support Vector Regression (SVR) is a part of support vector machine. Nowadays SVM is one of tools in data mining analysis which develop to solve the prediction and classification problem. It developed by structural risk minimization, which find the best class using hyperplane so that obtained minimum error. The major effort in kernel-based methods is the selection of an appropriate kernel function. Unsuitable kernel function may lead to significantly poor performance (Chapelle*etal.*, 2010). To compare the performance of different kernels, it can be implement three SVM kernels, namely linier, polynomial, and radial basis function (RBF).The one of advantages using SVM is able to obtain global optimum solution. It can be analyzed theoretically using concept from computational learning theory, and achieve good performance at the same time (Chen and Li, 2010). So that it has been successfully used to solve forecasting problems in many fields, such as financial time series forecasting, tourist arrival forecasting, traffic flow forecasting and many others. Many studies have been conducted for prediction using SVR such as rainfall forecasting (Hong, 2008), prediction of rainfall time series using modular using soft computing method (Wu and Chau 2013). Application of SVR with chaotic GASA algorithm to forecast Taiwanese 3G mobile phone demand (Li-Yueh, 2014). Modeling of daily evaporation (Goyal*et al.*, 2014) and short term load forecast (JinXing and Jian, 2014). Mohandes*et al.* (2004) shows applied SVM to predict wind speed. Their experimental results indicated that the SVM model out-performed multilayer neural network as measured by root mean square error. Cao (2003) presented a dynamic SVM model to solve non-stationary time series problem. They results indicated that dynamic SVM outperform

standard SVM in forecasting non-stationary time series problem.

Semarang city is one of the busiest city in Indonesia. Doe to its role as the capital city of Central Java, Semarang is known as having a relativity high-rate economic activities. The geographic of Semarang city bordered by the Java sea, thus whenever the rainfall is high, there could be flood at certain area. Therefore, prediction of rainfall is very important. The main aim of this research is prediction of weekly rainfallin Semarang city use support vector regression with quadratic loss function.

## II. MATERIAL AND METHOD

### 2.1 Support Vector Regression

SVM is initial applications focused on binary classification of test instances and pattern recognition. It is to construct a decision surface by mapping the input vectors into a high dimensional feature space (Wu and Chau 2013). With the introduction of Vapnik's $\varepsilon$ insensitive loss function, SVM has been extended to solve nonlinear regression estimation problem and has been successfully applied to solve prediction problem in many fields (Chapelle*etal.*, 2010).. Given training set data $\{x_i, y_i\}$ for $i = 1,2,\ldots,n$ , the aim of SVR is to induce a prediction which has good to predict. Where $x_i$ is the input vector; $y_i$ is the actual value and n is the total number of data, the SVR function is

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

where w represent the weight vector, b represent bias and $\varphi(x)$ is the feature of inputs. The set of vectors said to be optimally separated by hyperplane if it is separated without error and the distance between the closest vector to the hyperplane is maximal. From eq. (1) the coefficient are estimated by minimizing the following regularized risk function

$$R(f(x)) = \frac{1}{2}\|w\|^2 + \frac{C}{n}\sum_{i=1}^{n} L_\varepsilon(y_i, f(x_i)) \text{ where} \quad (2)$$

where

$$L_\varepsilon(y_i, f(x_i)) = \begin{cases} 0 & ;|y_i - f(x_i)| \le \varepsilon \\ |y_i - f(x_i)| - \varepsilon; & otherwise \end{cases}$$

$L_\varepsilon$ is called the $\varepsilon - insensitive$ loss function, $C$ and $\varepsilon$ are prescribed parameters. Concept of quadratic loss function is Eq. (2) can be transformed by minimize

$$R(w, \xi, \xi^*) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right)\right) \quad (3)$$

with constraint:

$w\varphi(x_i) + b - y_i \le \varepsilon + \xi_i^*$; $y_i - w\varphi(x_i) - b \le \varepsilon + \xi_i$ and

$\xi_i, \xi_i^* \ge 0$.

Eq. (3) is solved using Lagrangian form

$$L = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right)\right) - \sum_{i=1}^{n}\alpha_i\left[w\varphi(x_i) + b - y_i + \varepsilon + \xi_i\right]$$

$$-\sum_{i=1}^{n}\alpha_i^*\left[y_i - w\varphi(x_i) - b_i + \varepsilon + \xi_i^*\right] - \sum_{i=1}^{n}\left(\beta_i\xi_i + \beta_i^*\xi_i^*\right) \quad (4)$$

with applied Karush-Kuhn-Tuck condition, Eq. (4) can be result in Eq. (5)

$$Q(\alpha, \alpha^*) = -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\, K(x_i, x_j)$$

$$-\varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \quad (5)$$

with condition

$$\sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) = 0\,;\, 0 \le \alpha_i \le C\,;\, 0 \le \alpha_i^* \le C$$

where $K(x_i, x_j)$ called Kernel function. The selection parameters $C$ and $\varepsilon$ of SVR model is important accuracy of prediction. However, structural methods for confirming efficiently the selection of parameters efficiently are lacking. Optimal desired weight vector of the regression hyperplane is

$$w^* = \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)\varphi(x_i)$$

So that the regression function can be written

$$f(x_i, \alpha_i, \alpha_i^*) = \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right) K(x, x_i) + b$$

Kernel-based methods have been widely used for time series data analysis and forecasting. By using kernel function, SVR has been extended to solve nonlinear regression problems with linear method in an appropriate feature space (Li-Yueh, 2014). Thus performance of SVR is determined by the type of kernel function and the setting of kernel parameters. There are four types of kernel function, namely linear, polynomial, radial basis function and sigmoidal (Gunn, 1998).

The linear kernel is $K(x, x_i) = x_i^T x$;

polynomial kernel is $K(x, x_i) = \left(\gamma x_i^T x + r\right)^d$;

radial basis function is $K(x, x_i) = \exp\left\{-\gamma\left(\|x_i - x\|^2\right)\right\}$ and

sigmoidal kernel is $K(x, x_i) = \tanh\left(\gamma x_i^T x + r\right)$.

### 2.2 Method

In this study, focus on weekly rainfall data in Semarang city during 2009 until 2013. The amount of data is 256 data. The data is divided into two parts, among other training data (200 data) and testing data (56 data). The methodology of this research is:

Step 1 : Variable response is $y_t$ and variable predictor is $y_{t-1}$

Step 2 : Find kernel function (use linier, polynomial, and radial basis function

Step 3 : Find C parameter for each kernel function

Step 4 : Compute weighted vector and bias

Step 5 : Compute $\hat{y}$ of training data and testing data

Step 6 : Compute R-square of training data and testing data

Step 7 : Predict weekly rainfall using average of training data fit and testing data fit based on R-square maximum.

## III. RESULT AND DISCUSSION

Table 1 show that statistics descriptive of weekly rainfall in Semarang city. According to the Table. 1, average of weekly rainfall is 46.68 mm. It show that average of weekly

rainfall in Semarang is medium rainfall with maximum rainfall is 318.6 mm.

Table 1: Statistics Descriptive of weekly rainfall in Semarang city period 2009 to 2013

| | |
|---|---|
| Mean | 46.48 |
| Minimum | 0 |
| Maximum | 318.6 |
| Standard deviation | 52.08 |

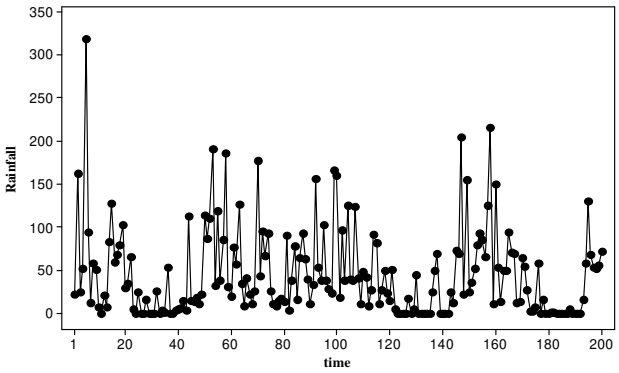


Figure 1: Scatter plot of weekly rainfall in Semarang city period 2008 to 2009

Figure 1 shows the scatter plot of actual weekly rainfall in Semarang city is non stationary. It configure pattern in certain periods.

Using linear, polynomial ($p$=1; p=2), and RBF kernel function for analysis. R-Square $\left(R^2\right)$ is

$$R^2 = \frac{\sum_{i=1}^{n} \varepsilon_i^2}{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2} \quad (8)$$

where $\varepsilon_i$ is residual of data. Then we find C parameter with trial and error for kernel function. It controls the trade-off between the regularization term and the training accuracy. Large values of C imply that more weight so that high error result.

Table 2: The weekly rainfall's R-Square using various Kernel function for training data

| Model | C | R-square | Model | C | R-square |
|---|---|---|---|---|---|
| Linear | 0.01 | 0.6951 | RBF=10 | 0.5 | 0.1961 |
| | 0.5 | 0.6953 | | 1 | 0.2347 |
| | 2 | 0.6835 | | 3 | 0.2649 |
| | 3 | 0.225 | | 10 | 0.2893 |
| polynomial p=1 | 0.01 | 0.5801 | | 30 | 0.2969 |
| | 0.5 | 0.4591 | | 50 | 0.2986 |
| | 2 | 0.456 | RBF=20 | 0.5 | 0.2374 |
| | 3 | 0.4564 | | 1 | 0.2676 |
| polynomial p=2 | 0.01 | 0.4783 | | 3 | 0.2933 |
| | 0.5 | 0.4476 | | 10 | 0.3176 |
| | **2** | **0.7161** | | 30 | 0.3243 |
| | 2.5 | 0.619 | | 50 | 0.3274 |

Table 2 shows the weekly rainfall's R-square using various kernel (linear, polynomial and RBF) for training data, the highest R-square is in polynomial (p=2) with C=2. For linear kernel function results around 60%, polynomial (p=1) kernel function around 40% and RBF kernel function around 20%. The optimal kernel function results R-square for training data is 71.61%. Then R-square of testing data is 76.64%. It's not too difference between R-square training data and testing data.

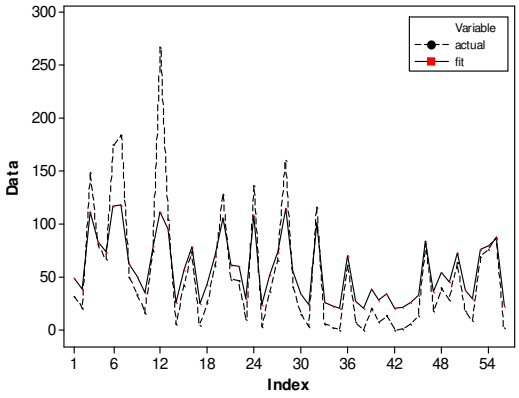


Figure 2: Scatter plot of actual and fit using SVR in testing data

Figure 2 shows SVR is able to break over fit of prediction.

Table 3: $\hat{y}_i$ for testing data

| period | fit | period | fit | period | fit | period | fit |
|---|---|---|---|---|---|---|---|
| 201 | 48.98 | 215 | 56.92 | 229 | 55.69 | 243 | 21.29 |
| 202 | 39.11 | 216 | 79.00 | 230 | 34.29 | 244 | 25.82 |
| 203 | 111.93 | 217 | 25.26 | 231 | 23.94 | 245 | 32.77 |
| 204 | 83.01 | 218 | 43.62 | 232 | 100.85 | 246 | 83.87 |
| 205 | 74.27 | 219 | 70.43 | 233 | 26.19 | 247 | 36.85 |
| 206 | 117.29 | 220 | 105.70 | 234 | 23.19 | 248 | 54.76 |
| 207 | 118.44 | 221 | 61.04 | 235 | 20.62 | 249 | 46.04 |
| 208 | 62.65 | 222 | 60.30 | 236 | 71.10 | 250 | 72.37 |
| 209 | 50.91 | 223 | 29.96 | 237 | 27.67 | 251 | 37.63 |
| 210 | 35.62 | 224 | 108.30 | 238 | 20.62 | 252 | 29.23 |
| 211 | 79.24 | 225 | 23.75 | 239 | 38.85 | 253 | 75.81 |
| 212 | 111.05 | 226 | 52.02 | 240 | 27.86 | 254 | 80.12 |
| 213 | 95.75 | 227 | 73.81 | 241 | 33.57 | 255 | 87.03 |
| 214 | 25.82 | 228 | 114.67 | 242 | 20.62 | 256 | 22.62 |

Table 3 shows that $\hat{y}_i$ testing data period 201 to 256 using polynomial kernel function for p=2 and C=2. Average of $\hat{y}_i$ for testing data is 56.96 mm and then average of $\hat{y}_i$ for training data 53.07 mm. Use average's fit of training data and testing data, for forecasting next period. Then average of rainfall in Semarang city is 55.02 mm. From analysis, weight vector of SVR $\|w\|^2$ is 425.51.

## IV. CONLUSIONS

The historical weekly rainfall data in Semarang city shows fluctuation trend, non stationary. Therefore, it can be result over prediction or under-prediction rainfall amount. This study introduced support vector regression (SVR) techniques to investigate its feasibility in prediction weekly rainfall amount. This method is able to break over-fitting prediction. The experimental results indicate that the SVR method has R-square 71.61% with polynomial (p=2) kernel function and C-parameter=2. Prediction the weekly rainfall in Semarang city for next period is 55.02 mm. It shows that the next period rainfall in Semarang city is medium rainfall.

ACKNOWLEDGMENT

The author would like to say thank you to laboratory of Statistics Department, Faculty of Science and Mathematics, Diponegoro University, for the supports during the research.

REFERENCES

Nawawi, G. 2001. *Pengantar Klimatologi Pertanian*. Modul Dasar Bidang Keahlian, Proyek pengembangan Sistem Standar Pengelolaan SMK, Direktorat Pendidikan Menengah Kejuruan, Departemen Pendidikan Nasional, Jakarta.

Cao, L. (2003). Support vector machine experts for time series modelling.*Neurocomputing*, Vol. 51, pp. 321-339.

Chapelle, O., Vapnik, V., Bousquet, O., and Mukherjee, S. (2002). Choosing multiple parameters for support vector machines.*Mach Learn*, vol. 46, no. 1-3, pp. 131-139.

Chen, F.L and Li, F.C. (2010).Combination of feature selection approaches with SVM in credit scoring.*Expert System Applications* 37, 4902-4909.

Goyal, *et. al*., "Modeling of Daily Pan Evaporation in Sub Tropical Climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS", *Expert System with Applications*, vol. 41, pp. 5267-5276, 2014.

Gunn, S. (1998).*Support Vector Machines for Classification*, University of Southamton.

Hong, W. C. 2008. Rainfall Forecasting by Technological Learning Models, International.*Applied Mathematics and Computation*, vol. 200, pp. 41-57.

JinXing, C., and JianZ.W. (2014).Short Term Load forecasting using kernel based support vector regression combination model.*Applied Energy* , vol. 132 pp 602-609.

Li-Yueh, C. (2014). Application of SVR with chaotic GASA algorithm to forecast Taiwanese 3G mobile phone demand.*Neuro computing 127 pp. 206-213*.

Mohandes, M.A., Halawani, T.O., Rehman, S., and Hussain,A.A. (2004). Support vector machines for wind speed prediction", *Renewable Energy*, vol. 29, no. 6, pp. 939-947.

Wu, C.L., andChau, K.W. 2013.Prediction of rainfall time series using modular soft Computing methods.*Engineering Applications of Artificial Intelligence*, vol. 26, pp. 997-1007.