Journal of Information Technology and Computer Science
Volume 5, Number 3, Desember 2020, pp. 235-246
Journal Homepage: www.jitecs.ub.ac.id

Comparison of Regression, Support Vector Regression (SVR), and SVR-Particle Swarm Optimization (PSO) for Rainfall Forecasting

Fendy Yulianto[1], Wayan Firdaus Mahmudy[2], Arief Andy Soebroto[3]

[1,2,3]Faculty of Computer Science, Brawijaya University
{fendy.yulianto37@gmail.com[1], wayanfm@ub.ac.id[2], ariefas@ub.ac.id[3]}

Received 18 January 2019; accepted 17 Desember 2020

Abstract. Rainfall is one of the factors that influence climate change in an area and is very difficult to predict, while rainfall information is very important for the community. Forecasting can be done using existing historical data with the help of mathematical computing in modeling. The Support Vector Regression (SVR) method is one method that can be used to predict non-linear rainfall data using a regression function. In calculations using the regression function, choosing the right SVR parameters is needed to produce forecasting with high accuracy. Particle Swarm Optimization (PSO) method is one method that can be used to optimize the parameters of the existing SVR method, so that it will produce SVR parameter values with high accuracy. Forecasting with rainfall data in Poncokusumo region using SVR-PSO has a performance evaluation value that refers to the value of Root Mean Square Error (RMSE). There are several Kernels that will be used in predicting rainfall using Regression, SVR, and SVR-PSO with Linear Kernels, Gaussian RBF Kernels, ANOVA RBF Kernels. The results of the performance evaluation values obtained by referring to the RMSE value for Regression is 56,098, SVR is 88,426, SVR-PSO method with Linear Kernel is 7.998, SVR-PSO method with Gaussian RBF Kernel is 27.172, and SVR-PSO method with ANOVA RBF Kernel is 2.193. Based on research that has been done, ANOVA RBF Kernel is a good Kernel on the SVR-PSO method for use in rainfall forecasting, because it has the best forecasting accuracy with the smallest RMSE value.

Keyword : Forecasting, SVR, Rainfall

1. Introduction

Indonesia is an archipelago that has a climate diversity where it is often referred to as the El-Nino and La-Nina phenomena [1]. From this phenomenon, Indonesia often experiences problems with the intensity of extreme rain that is affected by rainfall in conditions above normal and can cause floods and landslides. Rainfall is one element of the weather and includes a meteorological process which is quite difficult to predict [2].

Information about rainfall is very important, such as in agriculture where rainfall can determine which plants are good for planting under certain rainfall conditions [3]. Potato plants are one of the plants that affect the level of rainfall. Erratic rainfall is an obstacle that must be faced because it has a negative impact on the productivity of existing potatoes [4]. Due to the importance of rainfall factors, the researchers developed rainfall prediction methods that have a high degree of accuracy than predictions that have been made in the past [5].

In forecasting, various methods and models that use statistics or artificial intelligence have been used, and some are a combination of the two existing models. There is research on rainfall forecasting using Multiple Linear Regression (MLP) and several classical methods such as SVM, ANN, and several other methods [6]. From these results, it was found that the MLP method got better results than other methods in predicting rainfall, with a Mean Absolute Error (MAE) value of 0.0833.

There are other studies that also predict rainfall using Bayesian Regression, Support Vector Regression, and Wavelet Regression. Where in this study with the SVR method the smallest RMSE value is 108.71 in rainfall forecasting [7]. There are other studies that use the Long Short-Term Memory (LSTM) and LSTM-PSO methods in predicting rainfall. Particle Swarm Optimization (PSO) is used to optimize the parameters of the LSTM method and the results show that the LSTM-PSO method is better than the classical LSTM method. RMSE value obtained by LSTM-PSO was 0.149 while the classical LSTM was 0.166 [8].

SVR is the application of Support Vector Machine (SVM) in the case of Regression, a Regression approach that has been widely applied in solving forecasting problems. SVR builds a hyperplane in high dimensional space and can precisely distinguish objects from Kernel functions in linear or nonlinear data. SVR is a method that can overcome overfitting, so that it will produce good performance [9]. In SVR method itself there are several Kernels that can be used including linear, Gaussian, Polynomial and many other Kernels. The SVR method can also be added with the use of optimization and one of the optimizations is by using Particle Swarm Optimization (PSO). Where the addition of PSO optimization can improve the accuracy of the forecasting done [10].

PSO is an algorithm developed by Kennedy and Eberhart (1995), PSO itself is another technique of computational evolution [11]. The PSO algorithm itself is an optimization algorithm that is often used to solve optimization problems so that it is still often developed [12]. PSO processes search schemes using particle populations that are in accordance with individual use in genetic algorithms, each particle is equivalent to the solution of an existing problem [13]. As research conducted by [14] applied the Particle Swarm Optimization (PSO) for parameter optimization the SVR method to predict stock value of tata steel.

From several studies discussed above, Regression and SVR methods are good enough to make predictions. So that in this study a comparison will be made to find which method is better if you use the same data in forecasting. The methods used in this research are Regression, SVM, and SVM-PSO, so that in this study it is expected to know the strengths and weaknesses of each method in rainfall forecasting.

2. Methodology

The methodology in this research is a sequence of existing forecasting processes. The forecasting process is carried out based on the theory associated with the existing forecasting steps. In conducting this forecasting process refers to the problems that exist in the process of forecasting that will be carried out. In Fig. 1 is the steps of the forecasting process used in this study.

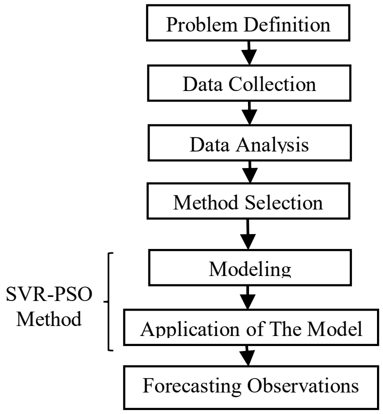

Fig. 1. Steps of the Forecasting Process

2.1. Data Collection

Data collection is used to obtain information about rainfall data that will be used in this study. The data used is the result of observations from the Meteorology, Climatology and Geophysics Agency for Climatology, Karangploso Malang Station. The following is the data specification used in this study.

- Rainfall data used comes from the Poncokusumo area in 2000-2020.
- Rainfall data ranges from ten days.
- Rainfall data in millimeters for ten days.

2.2. Method Selection

The forecasting process requires a method that can forecast well. There are two forecasting methods that can be used in conducting the forecasting process, namely quantitative forecasting methods and qualitative forecasting methods. Quantitative methods will be used in this study, because the data used are historical data with rainfall objects. In this study the SVR method is one of the quantitative methods chosen to conduct a test of forecasting the rainfall time in this study.

2.2.1. Support Vector Regression (SVR)

SVR is the application of Support Vector Machine (SVM) in the case of Regression, a Regression approach that has been widely applied in solving forecasting problems [15]. SVR builds a hyperplane in high dimensional space and can precisely distinguish objects from Kernel functions in linear or nonlinear data shown in equation 1. SVR is a method that can overcome overfitting, so that it will produce good performance [9].

$$f(x) = \sum_{j=1}^{m_1} w_j \phi_j(x) + b = (w.x) + b \tag{1}$$

Where:

$f(x)$ = predictive value
w = weight
φ = input space feature
x = data
b = bias value, or bias also represented byλ = lambda
m_1 = feature space

In order to obtain a decision function, the coefficient w_i and b must be estimated from the data. First, by defining ε-insentive loss function $L_\varepsilon(x, y, f(x))$, shown in equation 2.

$$L_\varepsilon(x, y, f(x)) = \begin{cases} |y - f(x)| - \varepsilon \; for |y - f(x)| \geq \varepsilon \\ 0 \quad Sebaliknya \end{cases} \tag{2}$$

Where:
$f(x)$ = predictive value
L = *flatness*
y = actual value
x = data
ε = loss value

The decline in SVR follows the principle of structural risk minimization rooted in the VC dimension theory. With the use of variables slack ξ_i and ξ_i^* can overcome the obstacles inability to convex optimization problems shown in equation 3.

$$\overset{min}{w \in \Re^N, \xi^{(')} \in \Re^{2N}}, b \in \Re \Phi(w, b, \xi_t, \xi_t^{'}) = \frac{1}{2}\|w\|^2 + C\left(\frac{1}{N}\sum_{1=1}^{N}(\xi_t + \xi_t^{'})\right)$$

$$\text{With the provision of} \begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, C \ge 0 \end{cases} \tag{3}$$

Where:
w = weight
$\in \Re$ = *vector input*
b = bias value, or bias also represented by λ = lambda
ε = loss value
ϕ = input space feature
C = complexity value
N = data value to -

Formulation of functions $\Phi(w, b, \xi_t, \xi_t^{'})$ very much in accordance with the principle of structural risk minimization. The first term $\frac{1}{2}\|w\|^2$, is a measure of flatness functions, minimizing what is related to maximizing the separation margin $\frac{2}{\|w\|}$, show maximum generalization ability.

Kernel trick function itself can handle cases with non-linear data, where there are several Kernel methods that are often used in the SVM method between the equations 4,5,6,7 and 8:

- Linear

$$K(x_1, x_2) = x_1.x_2 \tag{4}$$

- Polynomial

$$K(x_1, x_2) = (x_1.x_2 + c)^d \tag{5}$$

- Gaussian Radial Bassis Function (RBF)

$$K(x_1, x_2) = exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \tag{6}$$

- ANOVA Radial Bassis Function (RBF)

$$K(x_i, x_j) = \left(\sum_{i,j=1}^{N} exp\left(-\gamma_k (x_i - x_j)^2\right)\right)^d \tag{7}$$

- Invers Multiquadric

$$K(x_1, x_2) = \frac{1}{\sqrt{\|x_1 - x_2\|^2 + c^2}} \tag{8}$$

2.2.2. Particle Swarm Optimization (PSO)

PSO is an algorithm developed by Kennedy and Eberhart (1995), PSO itself is another technique of computational evolution. PSO algorithm itself is an optimization algorithm that is often used to solve optimization problems so that it is still often developed [12]. PSO processes the search scheme using particle populations according to individual use in the genetic algorithm shown in equation 9, each particle is equivalent to the solution to the existing problem.

$$X_i(t) = x_{i1}(t), x_{i2}(t), \dots, x_{iN}(t) \quad (9)$$

Where:
x = the best position ever passed
t = iteration value 1 to n
N = size of space dimensions
i = particle index

Each particle has the position and speed shown in equation 10.

$$V_i(t) = v_{i1}(t), v_{i2}(t), \dots, v_{iN}(t) \quad (10)$$

Where:
t = iteration value 1 to n
N = size of space dimensions
v = particle speed
i = particle index

Each particle iteration will approach the herd that has the best position from the others. Individuals in a herd will learn from experience in finding the best position [16]. Each particle has the speed shown in equation 11.

$$V_i(t) = V_i(t-1) + c_1 r_1\left(X_i^L - X_i(t-1)\right) + c_2 r_2\left(X^G - X_i(t-1)\right) \quad (11)$$

Where:
X_i^L = the best position ever passed
X^G = the best position the entire herd has ever passed
c = understanding position
v = particle speed
i = particle index
r = random value
t = iteration value 1 to n
X_i = the best position

Each particle has the best position in a herd shown in equation 12.

$$\mathrm{X_i(t) = V_i(t) + X_i(t-1)} \quad (12)$$

Where:
X_i = the best position
v = particle speed
i = particle index
t = iteration value 1 to n

In equation 10 $X_i^L = X_{i1}^L, X_{i2}^L, \dots, X_{iN}^L$ is the best position ever passed (local best), and $X^G = X_{i1}^G, X_{i2}^G, \dots, X_{iN}^G$ the best position that all the herds have ever gone through (global best).$c_1$is a process in understanding the best position by individual (learning rate), whereas c_2 is a process of understanding the best position of relationships between individuals. $r_1$and r_2 an ordinary random value can be initialized with 0 to 1 [12].

3. Calculation of the SVR-PSO Method

In the calculation process with the SVR method there are 3 Kernels to be used. Kernel in SVR method functions as the most important process in the SVR method itself. Kernels to be used include the Linear Kernel, Gaussian RBF Kernel, and ANOVA RBF Kernel. From the use of the three Kernels that are done, one Kernel will be selected which results in the best forecasting value from the other Kernels. formulations and calculations are needed according to the theory. In Fig. 2 is the flow of the calculation process using the SVR-PSO method in forecasting the time of rainfall.

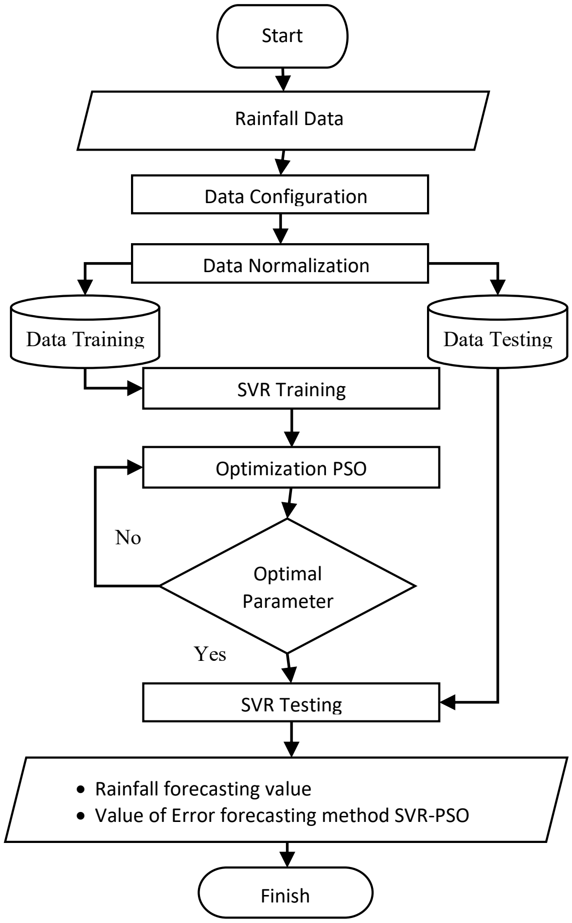

Fig. 2. Flowchart SVR-PSO method

3.1. Test SVR-PSO Method

In this study, the implementation of a prototype-based software system that can forecast rainfall time using SVR-PSO method. So that it can be tested SVR-PSO method in forecasting the forecasting of the time of rainfall for the system that has been implemented. The following are the results of tests that have been carried out on the SVR-PSO method.

3.1.1. SVR Parameter Limit Test

The test of SVR parameter limits is intended to limit the particle dimensions in finding a solution so that it can produce a combination of SVR parameters that are optimal in the training process. Testing the SVR parameter limits include gamma (γ), lambda (λ), complexity (c), and epsilon (ε). The testing of SVR parameter limits is done in the SVR training process using rainfall data for ten days in a month from January 2000 to January 2020.

3.1.2. Gamma Parameter Limit Test (γ)

The limit of test values used in the Y parameter test consists of 0.00001 - 0.0001, 0.00001 - 0.001, 0.00001 - 0.01, 0.0001 - 0.001, 0.0001 - 0.01, 0.0001 - 0.1, 0.001 - 0.01, 0.001 - 0.1, and 0.001 - 10,0001 - 100. Each test was conducted five times. In Fig. 3 is the result of the test of the gamma parameter limit (γ) represented in graphical form.

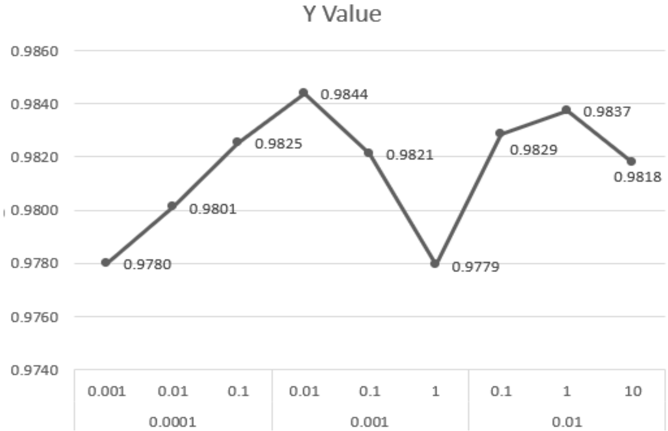

Fig. 3. Graph of gamma (γ) parameter limits test results

3.1.3. Lambda Parameter Limit Test (λ)

The limit of the test value used in testing the parameter λ consists of 0.0001 - 0.001, 0.0001 - 0.01, 0.0001 - 0.1, 0.001 - 0.01, 0.001 - 0.1, 0.001 - 1, 0.01 - 0.1, 0.01 - 1, and 0.01 - 10. Each test trying five times. In Fig. 4 is the result of testing the boundary of the lambda parameter (λ) represented in graphical form.

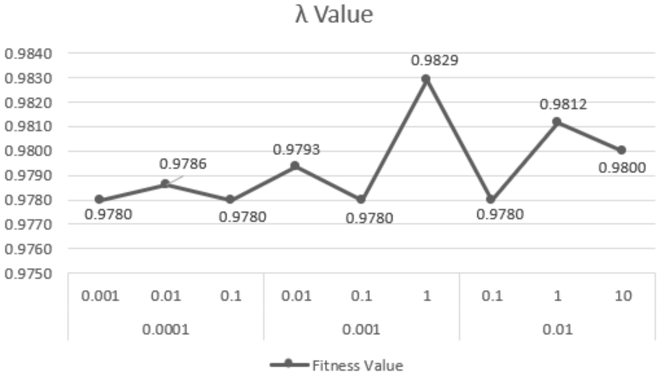

Fig. 4. Graph of lambda (λ) parameter limits test results

3.1.4. Complexity Parameter Limits Test (c)

The limit of test values used in parameter C testing consists of 1-10, 1 - 100, 1 -

1000, 10 - 100, 10 - 1000 and 100 - 1000. Each test is done five times. In Fig. 5 is the result of testing the boundary of the Complexity parameter (c) represented in graphical form.

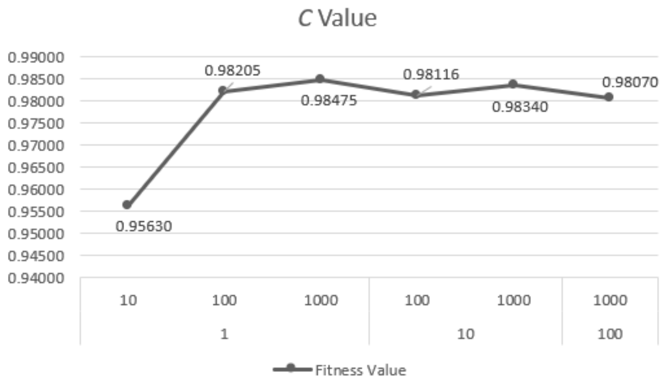

Fig. 5. Graph of complexity (c) parameter limits test results

3.1.5. Epsilon Parameter Limit Test (Ɛ)

The limits of the test values used in parameter testing e consist of 0.000001 - 0.00001, 0.000001 - 0.0001, 0.000001 - 0.001, 0.00001 - 0.0001, 0.00001 - 0.001, 0.00001 - 0.01, 0.0001 - 0.001, 0.0001 - 0.01, and 0.0001 - 0.1. Each test was conducted five times. in Fig. 6 is the result of testing the epsilon parameter limit (() represented in graphical form.

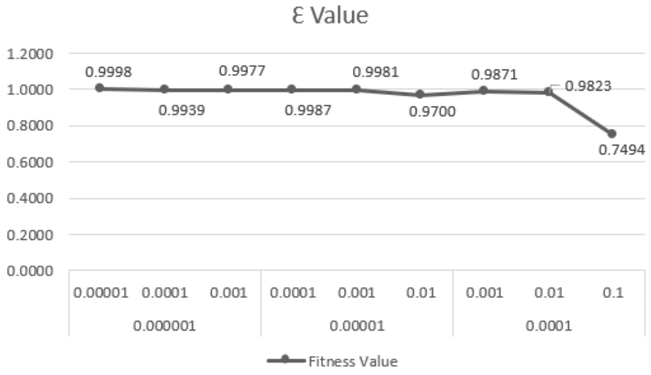

Fig 6. Graph of epsilon (Ɛ) parameter limit test results

3.1.6. Test the Number of SVR Iterations

To find out the right iteration in carrying out calculations using the SVR method, it is necessary to test existing iterations. The following is the SVR parameter limit used in carrying out the iteration testing process:

a. Y parameter limit: 0.001– 0.01
b. Limits of parameters λ: 0.01 - 1
c. Parameter limit C: 1 - 1000
d. Parameter limits ε: 0.000001 - 0.00001

In Fig. 7 is the result of the test of the number of SVR iterations represented in graphical form.

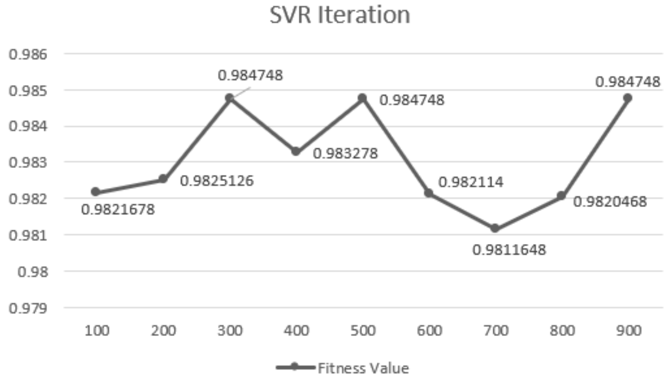

Fig. 7. Graph of SVR iteration test results

3.2. Test PSO Particles

Testing the number of particles in the PSO method is used to find out how many particles are needed in the process of optimizing the SVR method parameters to get optimal results. Test the number of particles from the PSO method with the number of particles 5, 50, 100, 200, 250, 400, 550, 700, 850 and 1000 in the SVR training process uses rainfall data for ten days every month from January 2000 to December 2018. Each trial was conducted five times, in Fig. 8 is the result of testing the number of PSO particles represented in graphical form.

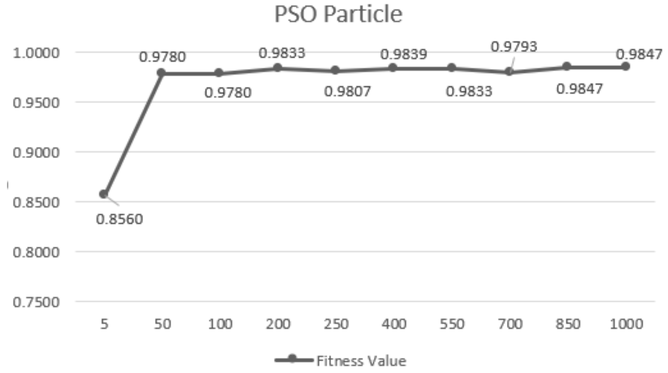

Fig. 8. Graph of PSO particle test results

3.2.1. Test the Number of PSO Iterations

Testing the number of iterations in the PSO method is used to find out how many iterations are needed in the process of optimizing the SVR method parameters to get optimal results. The trial of the number of iterations from the PSO method with the number of iterations 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 in the SVR training process uses rainfall data for ten days every month from January 2000 to December 2018. Each trial was conducted five times, in Fig. 9 is the result of testing the number of PSO iterations represented in graphical form.

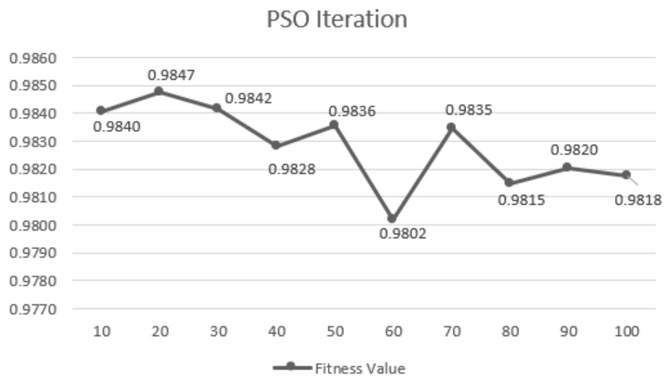

Fig. 9. Graph of PSO iteration test results

3.3. Rainfall Forecasting Trials

Forecasting trials conducted in this study use data for ten days and every month there are three rainfall data from January 2000 to January 2020. This forecasting trial consists of 7 methods namely multiple Linear Regression, SVR (Linear), SVR (Gaussian RBF), SVR (ANOVA RBF), SVR-PSO (Linear), SVR-PSO (Gaussian RBF), and SVR-PSO (ANOVA RBF). The following are the results of forecasting in table form which can be seen in Table 1.

Table 1 Results of regression formation f (x) training data and test data

	RMSE	**Fitness**	**Forecasting Data 1**	**Forecasting Data 2**	**Forecasting Data 3**
Actual			90	63	119
Regression	56.098	0.013	123.877	52.336	42.633
SVR Linear Kernel	88.426	0.018	135.321	52.597	67.626
SVR Gaussian Kernel	22.837	0.031	86.122	54.894	83.769
SVR ANOVA Kernel	12.982	0.079	71.231	56.743	92.226
SVR-PSO Linear Kernel	7.998	0.134	79.009	59.942	109.331
SVR-PSO Gaussian Kernel	27.172	0.027	77.55791	67.934	87.209
SVR-PSO ANOVA Kernel	2.193	0.323	87.321	62.933	118.383

Based on the results of the forecasting trials carried out using 7 different methods, it was found that the RMSE value was cursed on SVR-PSO method (ANOVA RBF). RMSE value obtained from SVR-PSO method (ANOVA RBF) is 2.193 using forecasting data from the years 2000-2020. The data used in the forecasting of rainfall time is ten days data in January 2000 - 2017 with the selected test data being in 2018, 2019, and 2020. In Figure 10 is the result of the trial of the forecasting of the time of ten days of rainfall in January with forecasting targets in 2018, 2019, and 2020 which are represented in graphical form.

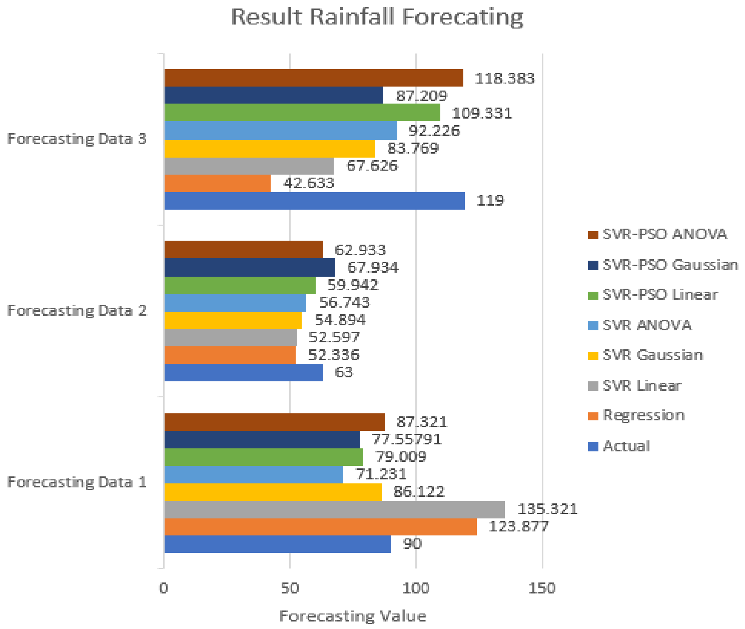

Fig. 10. Graph of ten days rainfall forecasting results for January 2018, 2019, and 2020 with 7 methods

4. Conclusion

The results of the study of rainfall forecasting by comparing the 3 methods can be concluded that the SVR-PSO gets the best results from other methods in terms of forecasting rainfall based on performance evaluation which refers to the RMSE value. The RMSE value obtained is the result of the comparison actual data with the existing forecasting data, smaller RMSE value obtained the better forecasting results obtained. Based on the results of tests that have been carried out using SVR-PSO method with Linear Kernel, Gaussian RBF Kernel, and ANOVA RBF Kernel, RMSE values are different for each Kernel. RMSE results were obtained in the process of forecasting the ten days of rainfall in January 2019 using Linear Kernel of 7.998, Gaussian RBF Kernel of 27.172, and ANOVA RBF Kernel of 2.193.

References

[1] C. H. Fajri, H. Siregar dan Sahara, "Impact of Climate Change on Food Price in The Affected Provinces of EL NINO and LA NINA Phenomenon: case of Indonesia," *International Journal of Food and Agricultural Economics,* vol. 7, no. 2147-8988, pp. 329-339, 2019.

[2] J. Refonaa, M. Lakshmi, R. Abbas dan M. Raziullha, "Rainfall Prediction using Regression Model," *International Journal of Recent Technology and Engineering (IJRTE),* vol. 8, no. 2S3, pp. 2277-3878, 2019.

[3] Z. T. Geneti, "Review on the Effect of Moisture or Rain Fall on Crop Production," *Civil and Environmental Research ,* vol. 11, no. 2224-5790 , 2019.

[4] I. Wahyuni, E. F. P. Adipraja dan W. F. Mahmudy, “Determining Growing Season of Potatoes Based on Rainfall Prediction Result Using System Dynamics,” *Indonesian Journal of Electrical Engineering and Informatics (IJEEI),* vol. 6, no. 2089-3272, pp. 210-216, 2018.

[5] M. Mohammed, R. Kolapalli, N. Golla dan S. S. Maturi, “Prediction Of Rainfall Using Machine Learning Techniques,” *International Journal of Scientific& Technology Research,* vol. 9, no. 1, pp. 2277-8616, 2020.

[6] N. Gnanasankaran dan E. Ramaraj, “A Multiple Linear Regression Model To Predict Rainfall Using Indian Meteorological Data,” *International Journal of Advanced Science and Technology,* vol. 29, no. 9, pp. 746-758, 2020.

[7] A. Sharma dan M. K. Goyal, “A Comparison of Three Soft Computing Techniques, Bayesian Regression, Support Vector Regression, and Wavelet Regression, for Monthly Rainfall Forecast,” *Journal Intelligent System,* vol. 4, no. 26, p. 641–655, 2017.

[8] S. Geetha, “Improving the Accuracy of Rainfall Prediction using Optimized LSTM Model,” *International Journal of Recent Technology and Engineering (IJRTE),* vol. 7, no. 6S, pp. 2277-3878, 2019.

[9] A. K. Lagat, A. G. Waititu dan A. K. Wanjoya, “Support Vector Regression and Artificial Neural Network Approaches: Case of Economic Growth in East Africa Community,” *American Journal of Theoretical and Applied Statistics,* vol. 7, no. 2326-8999, pp. 67-79, 2018.

[10] E. M. Priliani, A. T. Putra dan M. A. Muslim, “Forecasting Inflation Rate Using Support Vector Regression (SVR) Based Weight Attribute Particle Swarm Optimization (WAPSO),” *Scientific Journal of Informatics,* vol. 5, no. 2407-7658, 2018.

[11] O. F. Aje dan A. A. Josephat, “The particle swarm optimization (PSO) algorithm application – A review,” *Global Journal of Engineering and Technology Advances,* vol. 3, p. 001–006, 2020.

[12] S. A. Kadam dan D. M. S. Chavan, “Performance Of Particle Swarm Optimization For Reducing System Delay,” *International Journal of Scientific & Technology Research,* vol. 8, no. 11, pp. 2277-8616, 2019.

[13] Y. Sun dan Y. Gao, “An Efficient Modified Particle Swarm Optimization Algorithm for Solving Mixed-Integer Nonlinear Programming Problems,” *International Journal of Computational Intelligence Systems,* vol. 12, no. 1875-6891, pp. 530-543, 2019.

[14] M. Siddique, S. Mohanty dan D. Panda, “A Hybrid Forecasting Model for Prediction of Stock Value of Tata Steel Using Support Vector Regression and Particle Swarm Ooptimization,” *International Journal of Pure and Applied Mathematics,* vol. 119, pp. 1719-1727, 2018.

[15] S. Vijayakumar dan S. Wu, “Sequential Support Vector Classifiers and Regression,” *IIA/SOCO,* pp. 610-619, June 1999.

[16] J. Kennedy dan R. Eberhart, Particle Swarm Optimization, Indianapolis, IN 46202-5160 : Institute of Electrical and Electronics Engineers (IEEE), 1995.