Forecasting the Length of the Rainy Season Using Time Delay Neural Network

Agus Buono, Muhammad Asyhar Agmalaro, and Amalia Fitranty Almira
Department of Computer Science
Faculty of Mathematics and Natural Sciences, Bogor Agricultural University
Email: pudesha@gmail.com, agmalaro@gmail.com, lia.ilkom47@gmail.com

Abstract—**Indonesia has abundant natural resources in agriculture. Good agricultural results can be obtained by determining a good growing season plan. One of important factors which determines the successful of crop is the length of the rainy season. The length of the rainy season is dynamic and difficult to be controlled. Therefore the planning of the growing season becomes inaccurate and cause crop failures. This research aims to develop a model to predict the length of the rainy season using time delay neural network (TDNN). Observational data used in this research is the length of rainy season from three weather and climate stations of the Pacitan region from 1982/1983 to 2011/2012. Predictor data used in this reserach is sea surface temperature (SST) derived from the region of Nino 1+2, Nino 3, Nino 4, and Nino 3.4 from 1982 to 2011. Model with the best accuracy was obtained by Pringkuku station with RMSE of 1.97 with parameters of delay [0 1 2 3], learning rate 0.1, 40 hidden neurons, and predictors of Nino 3 and R-squared of 0.82 with parameters of delay [0 1], learning rate 0.3, 5 hidden neurons, and predictors of Nino 3.**

I. INTRODUCTION

Indonesia has abundant natural resources in agriculture. The quality of the agricultural products is influenced by a good planning of growing season. One important factor in a good planning of growing season is the length of the rainy season which is difficult to predict. Indonesia passed by west and east monsoon wind so Indonesia have two seasons, rainy and dry season. Season in Indonesia is also influenced by global phenomena as El Nino and La Nina [1]. In tropical areas, El Nino and La Nina lead to a shift in rainfall patterns, change in the amount of rainfall, and change the air temperature [2]. This leads to the difficulty in predict the length of the rainy season so that the planning of growing season becomes less precise and have an impact on crop failure.

The length of the rainy season is greatly affect rice production, especially in the second growing season. If the rainy season is short, then the chances of drought during the second growing increase and can cause the crop failure [3]. Information regarding the length of the rainy season is very useful for the parties involved in the good planning of growing season as much as possible in order to avoid crop failure and minimal losses. This study aims to solve the problem by building a good model in predicting the length of the rainy season.

Prediction of the length of the rainy season in this study is using time delay neural network (TDNN) with SST as the predictor. TDNN is able to capture the diverse characteristics of the data [4] so it is suitable for the length of the rainy season data that diverse and uncertain. This method has extraction layer using a sliding window on the input layer so it is dynamic [4].

Predictors used in this study are SST which is one of the global phenomenon that affects some variable rainfall and one of them is the length of the rainy season [5] and it was showed that there is strong correlation between SST with rainfall in Indonesia [6]. The length of the rainy season data that used in this study is from Pacitan area. Based on data Pacitan from 1982 to 2009 indicated that about 90% of drought occured in the dry season (May, June, July, and August) and the rest occurred in the early rainy season (November and December), it is interesting because theoretically rain should be abundance in this period, this indicates that there is a unique pattern in this data [7].

II. RESEARCH METHODS

A. Data Collection

Data that used in the study were SST as a predictor and the length of the rainy season as an observation. SST data is obtained from the National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Agriculture from the region of Nino 1 +2, Nino 3, Nino 3.4 and Nino 4 in each month from 1982 to 2011. The length of the rainy season data on climate and weather stations in each Pacitan region,

ISBN : 978-979-1421-225

obtained from the Center for Climate Risk and Opportunity Management in Southeast Asia Pacific (CCROM SEAP) Bogor Agricultural University from 1982/1983 to 2011/2012, derived from climate and weather station Arjosari, Kebon Agung, and Pringkuku. Unit that used for the length of the rainy season is dasarian that is grouping in ten days time so that in a year divided into 36 [3]. Dasarian of the length of the rainy season is obtained by calculating how many dasarian from early rainy season until end of the rainy season. The beginning of the rainy season is dasarian that has rainfall greater than or equal to 50 millimeters, followed by two subsequent dasarian [8].

B. Selection of Data

The selection of data is done by a simple correlation analysis of the length of the rainy season data and the SST data to get months that used as predictors. SST data for each month in each Nino will be calculated correlation value with the length of the rainy season data at each station. Correlation values were calculated using the linear correlation coefficient, it was a linear relationship between two random variables X and Y is denoted by r [9]. Correlation value equation is as follows:

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2\right]}}$$

where x_i is variable 1, y_i is variable 2, n is total data.

Every month will have correlation value and will select the month that correlated significantly with the length of the rainy season at 10% significance level by comparing the correlation values at 10% significance level with the correlation value of the length of the rainy season and SST every month. Correlation values at 10% significance level obtained by the equation:

$$r = \frac{t}{\sqrt{(n-2)+t^2}}$$

where t is t table value at 10% significance and n is total data. Selected months as predictors are the months that have an influence on the length of the rainy season one period forward [3].

C. Time Delay Neural Network Process

The data will be divided into 3 groups of data with training data as much as 2/3 and test data as much as 1/3 using 3-fold cross validation. In k-fold cross validation the data set is divided into k-subsets which are independent with the same size where k-1 subset is used as training data and 1 subset used as test data [10]. The training process will be carried out on the training data using a learning algorithm TDNN. TDNN consists of two main parts they were the extraction and grouping section. Before go to the grouping section, on the extraction section data will be scanned using a sliding window with a predetermined size, here in after referred as delay, after that, the data gets into the grouping section where the process is similar with artificial neural network in general [4]. Figure 1 shows an illustration of the TDNN architecture with 1 variable input, delay [0 1], and the 2 hidden neurons.

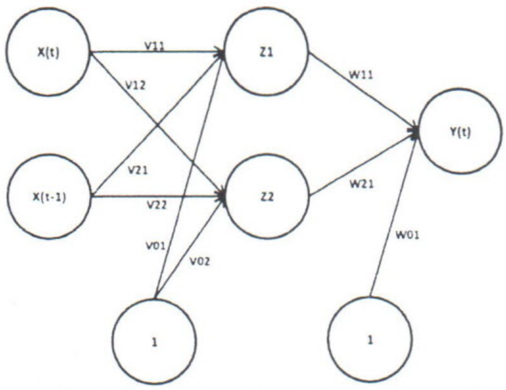

Fig. 1. Architecture of time delay neural network

The learning process is done to get the TDNN architecture that has been weighted to be used to predict the length of the rainy season using test data. Architecture used in this study using one hidden layer with 5, 10, 20, and 40 hidden neurons. Delay that used in this study were [0 1], [0 1 2], and [0, 1, 2, 3]. Activation function in the hidden layer is log-sigmoid while the output layer is linear. Learning rate used in this study were 0.3, 0.1, and 0.01. The training algorithm was resilient backpropagation that used in this study.

D. Analysis and Evaluation

Analysis and evaluation of the test data results is done by calculating the root mean square error (RMSE) and coefficient of determination (R^2 or R-squared). R-squared is used to measure how well the regression line formed by the estimated value with the actual value, the better if closer to 1. R-squared is defined as follows:

$$R^2 = \frac{\left[\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})\right]^2}{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where y_i is actual value and $\hat{y}_i$ is estimated value.

RMSE is shows the amount of deviation between the estimated value and the actual value, the better if it is close to 0. RMSE is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(X_t - F_t)^2}{n}}$$

where X_t is actual value at time of t and F_t is estimated value at time of t.

ISBN : 978-979-1421-225

III. RESULTS

A. Predictors Selection

Figure 2 shows the correlation between SST value every month in each Nino with the length of the rainy eason Arjosari, the red line is the correlation value at 10% significance level. Predictors used in Arjosari station from Nino 1 +2 are September, October, November, and December. Predictors from Nino 3 are August, September, October, November, and December. Predictors from Nino 4 are July, August, September, October, and November. Predictors from Nino 3.4 are July, August, September, October, November, and December.

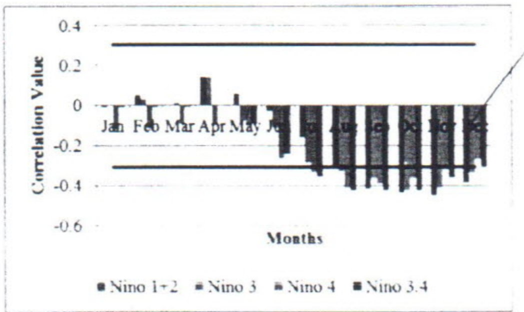

Fig. 2. Correlation value of Arjosari station

Figure 3 shows the correlation between SST value every month in each Nino with the length of the rainy eason Kebon Agung, the red line is the correlation value at 10% significance level. Predictors used in Kebon Agung station from Nino 1 +2 are April and November. Predictors from Nino 3 are March, April, November, and December. Predictors from Nino 4 are July, August, September, October, November, and December. Predictors from Nino 3.4 are August, September, October, November, and December.

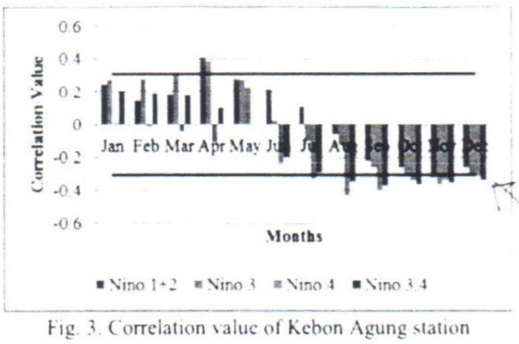

Fig. 3. Correlation value of Kebon Agung station

Figure 4 shows the correlation between SST value every month in each Nino with the length of the rainy eason Pringkuku, the red line is the correlation value at 10% significance level. Predictors used Pringkuku station from Nino 1 +2 are August, September, October, November, and December. Predictors from Nino 3 are June, July, August, September, October, November, and December. Predictors from Nino 4 are April, June, July, August, September, October, November, and December. Predictors from Nino 3.4 are June, July, August, September, October, November, and December.

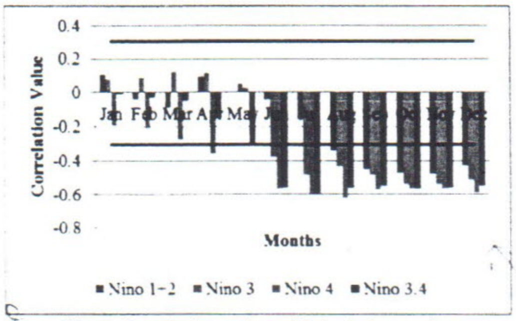

Fig. 4. Correlation value of Pringkuku station

B. Best Parameters

At Arjosari station, highest R-squared value is 0.75 with RMSE value of 2.40, this RMSE value is not the lowest value generated. Highest R-squared obtained with parameters of delay [0 1 2 3], learning rate 0.3, 5 hidden neurons, and predictors from Nino 1 +2. R-squared of 0.75 indicates that 75% of variability in the values of the prediction results can be explained by a linear relationship with the actual values. RMSE of 2.40 indicates that the average error for predicting the length of the rainy season is 2.40 dasarian or about 24 days. Figure 5 shows the scatter diagram between the predictions data and the observational data from Arjosari station with a positive correlation value of 0.87, which means between the observational data and the prediction data has an unidirectional relationship. Figure 6 shows a prediction graph Arjosari station with highest R-squared.

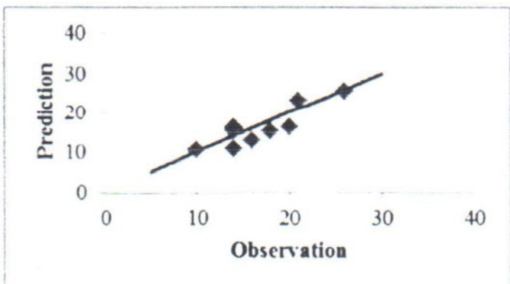

Fig. 5. Scatter diagram Arjosari with highest R-squared

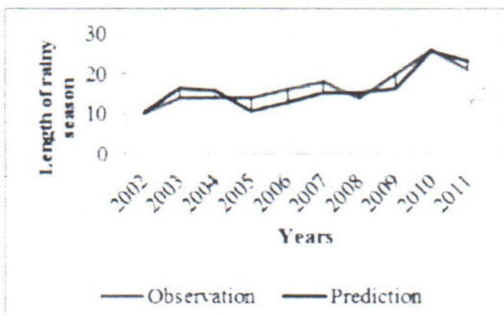

Fig. 6. Arjosari station with highest R-squared

ISBN : 978-979-1421-225

The lowest RMSE values at Arjosari station is 2.09 with R-squared value of 0.41, this RMSE value is lower than the previous value of 0.31 and R-squared value is lower than the previous value of 0.34. The lowest RMSE values obtained with parameters of delay [0 1 2 3], learning rate 0.1, 40 hidden neurons, and predictors from Nino 4. Figure 7 shows the scatter diagram between the prediction data and the observation data from Arjosari station with a positive correlation value of 0.64. Figure 8 shows a prediction graph Arjosari station with lowest RMSE.

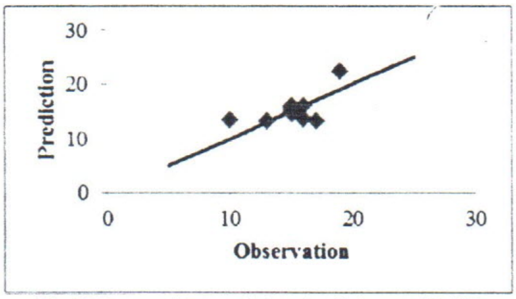

Fig. 7. Scatter diagram Arjosari with lowest RMSE

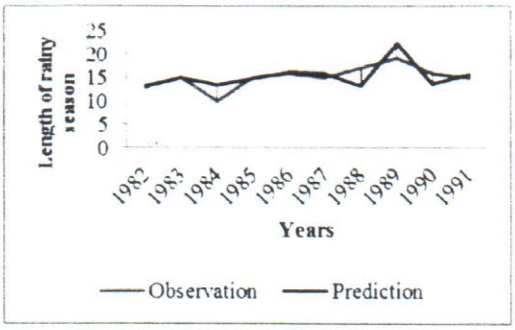

Fig. 8. Arjosari station with lowest RMSE

At Kebon Agung station, highest R-squared value is 0.74 with RMSE value of 5.86, this RMSE value is not the lowest value generated. Highest R-squared obtained with parameters of delay [0 1], learning rate 0.3, 20 hidden neurons, and predictors from Nino 3. Figure 9 shows the scatter diagram between the prediction data and the observation data from Kebon Agung station with a positive correlation value of 0.86. Figure 10 shows a prediction graph Kebun Agung station with highest R-squared.

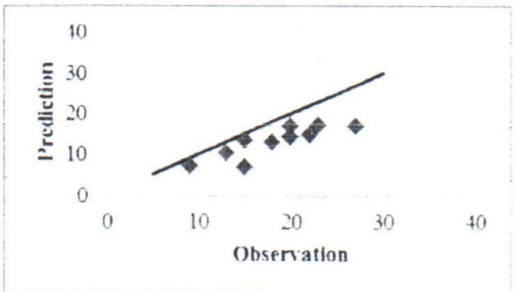

Fig. 9. Scatter diagram Kebon Agung with highest R-squared

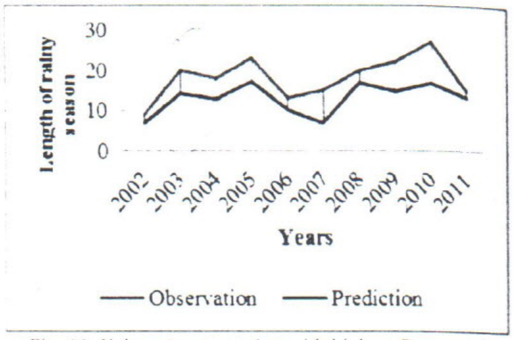

Fig. 10. Kebon Agung station with highest R-squared

The lowest RMSE values at Kebon Agung station is 3.47 with R-squared value of 0.05, this RMSE value is lower than the previous value of 2.39 and R-squared value is lower than the previous value of 0.69. The lowest RMSE values obtained with parameters of delay [0 1 2], learning rate 0.01, 5 hidden neurons, and predictors from Nino 1+2. Figure 11 shows the scatter diagram between the prediction data and the observation data from Kebon Agung station with a positive correlation value of 0.22. Figure 12 shows a prediction graph Kebon Agung station with lowest RMSE.

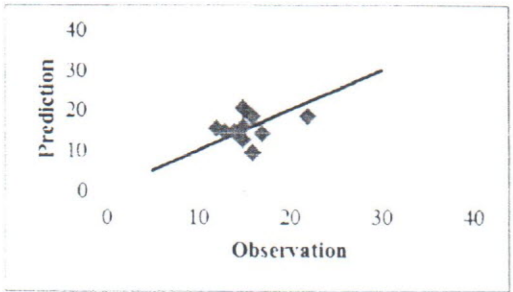

Fig. 11. Scatter diagram Kebon Agung with lowest RMSE

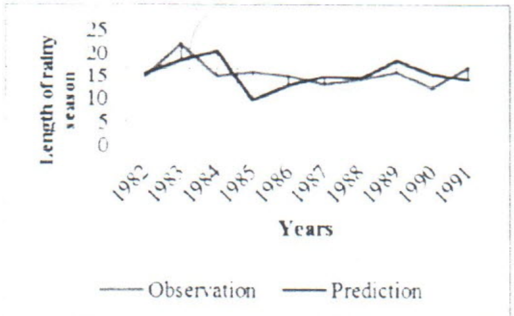

Fig. 12. Kebon Agung station with lowest RMSE

At Pringkuku station, highest R-squared value is 0.82 with RMSE value of 2.41, this RMSE value is not the lowest value generated. Highest R-squared obtained with parameters of delay [0 1], learning rate 0.3, 5 hidden neurons, and predictors from Nino 3. Figure 13 shows the scatter diagram between the prediction data and the observation data from Pringkuku station with a positive correlation value of 0.90. Figure 14 shows a prediction graph Pringkuku station with highest R-squared.

ISBN : 978-979-1421-225

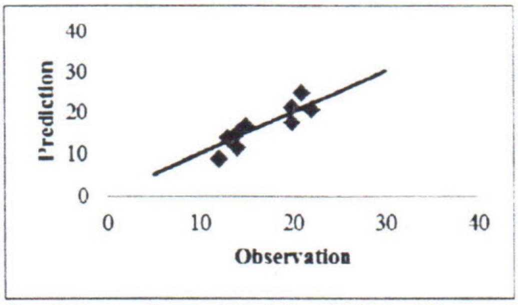

Fig. 13. Scatter diagram Pringkuku with highest R-squared

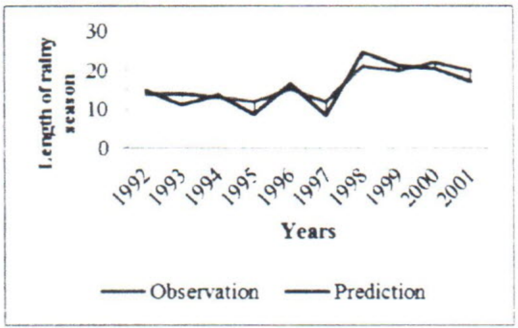

Fig. 14. Pringkuku station with highest R-squared

The lowest RMSE values at Pringkuku station is 1.97 with R-squared value of 0.77, this RMSE value is lower than the previous value of 0.44 and R-squared value is lower than the previous value of 0.05. The lowest RMSE values obtained with parameters of delay [0 1 2 3], learning rate 0.1, 40 hidden neurons, and predictors from Nino 3. Figure 15 shows the scatter diagram between the prediction data and the observation data from Pringkuku station with a positive correlation value of 0.88. Figure 16 shows a prediction graph Pringkuku station with lowest RMSE.

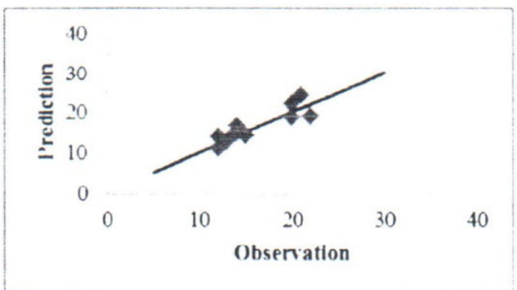

Fig. 15. Scatter diagram Pringkuku with lowest RMSE

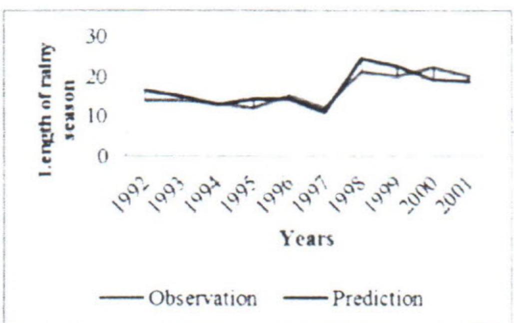

Fig. 16. Pringkuku station with lowest RMSE

C. *Comparison of All Stations*

Architecture to predict the length of the rainy season is good if the RMSE values close to 0 and the R-squared value close to 1. Figure 17 shows the RMSE and R-squared value at each station. Pringkuku station has the lowest RMSE value compared to the other stations, while Kebon Agung station has the highest RMSE value compared to the other stations. Also, Pringkuku station has the highest R-squared value compared to the other stations, while Kebon Agung station has the lowest R-squared value compared to the other stations. Overall, Pringkuku station has the best architecture to predict the length of the rainy season.

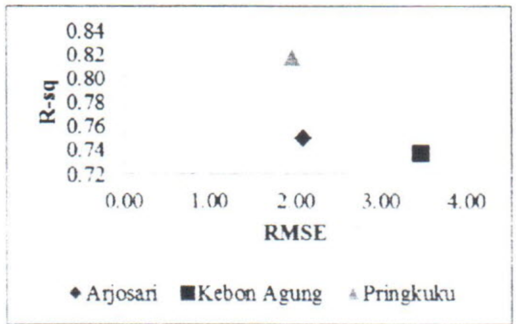

Fig. 17. RMSE and R-squared of each stations

D. *Distribution of R-squared and RMSE*

Parameters of each station that produces the best architecture retrained for 10 times and the result is the value of RMSE and R-squared that different for each training. Figure 18 shows the distribution of R-squared value from each station after retrained for 10 times. Distribution of the R-squared values from Arjosari station is asymmetric and skewed to the left with the median value closer to the third quartile. Distribution of the R-squared values from Kebon Agung station is also asymmetric and skewed to right with the median value closer to the first quartile and there is an outlier value above the maximum limit of the boxplot. Distribution of the R-squared values from Pringkuku station is also asymmetric and skewed to the left with the median value closer to the third quartile.

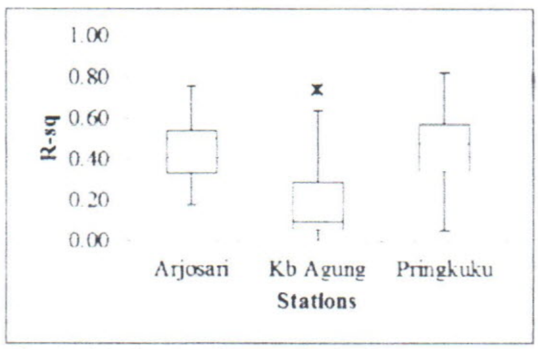

Fig. 18. Distribution of R-squared value from each stations

ISBN : 978-979-1421-225

Figure 19 shows the distribution of RMSE value from each station after retrained for 10 times. Distribution of the RMSE values from Arjosari station is asymmetric and skewed to the left with the median value a little closer to the third quartile. Distribution of the RMSE values from Kebon Agung station is also asymmetric and skewed to the right with a median value closer to the first quartile and there are outlier values below the minimum and above the maximum limit of the boxplot. Distribution of the RMSE values from Pringkuku station is also asymmetric and skewed to the left with the median value closer to the third quartile and there is an outlier value above the maximum limit of the boxplot.

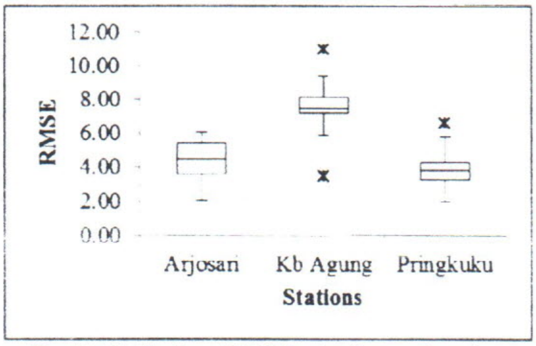

Fig. 19. Distribution of RMSE value from each stations

Prediction of the length of the rainy season results is different for each repetition despite using the same parameters, resulting different RMSE and R-squared values for each training. The best results are not always produced by the same parameters that the best results can be generated by different parameters with different weights initialization. This is due to the parameters used in this study are set manually and less thorough search parameters so that the optimum value can not be achieved with the use of these parameters. Therefore, search parameter values more thoroughly is required to obtain the optimum results.

IV. Conclusion

Conclusions from these results that Arjosari station obtained the best R-squared value of 0.75 with delay [0 1 2 3], learning rate 0.3, 5 hidden neurons, and predictors of Nino 1+2, and the best RMSE value of 2.09 with delay [0 1 2 3], learning rate 0.1, 40 hidden neurons, and predictors of Nino 4. Kebun Agung station obtained the best R-squared value of 0.74 with delay [0 1], learning rate 0.3, 20 hidden neurons, and predictors of Nino 3, and the best RMSE value of 3.47 with delay [0 1 2], learning rate 0.01, 5 hidden neurons, and predictors of Nino 1+2. Pringkuku station obtained the best R-squared value of 0.82 with delay [0 1], learning rate 0.3, 5 hidden neurons, and predictors of Nino 3, and the best RMSE value of 1.97 with delay [0 1 2 3], learning rate 0.1, 40 hidden neurons, and predictors of Nino 3.

References

[1] M. M. Said, "Peramalan panjang musim hujan menggunakan jaringan syaraf tiruan *resilient backpropagation*," Bogor: Bogor Agricultural University, 2011.

[2] B. Irawan, "Fenomena anomali iklim El Nino dan La Nina: kecenderungan jangka panjang dan pengaruhnya terhadap produksi pangan," *Forum Penelitian Agro Ekonomi*, vol. 24, no. 1, pp. 28-45, 2006.

[3] A. Buono, M. Mukhlis, A. Faqih, and R. Boer, "Pemodelan jaringan syaraf tiruan untuk prediksi panjang musim hujan berdasar sea surface temperature," *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, Jun. 2012.

[4] K. Benmahdjoub, Z. Ameur, and M. Boulifa, "Forecasting of rainfall using time *delay* neural network in Tizi-Ouzou (Algeria)," *Energy Procedia*, vol. 36, pp. 1138-1146, 2013.

[5] F. Lo, M. C. Wheeler, H. Meinke, and A. Donald, "Probabilistic forecasts of the onset of the North Australian wet season," *Monthly Weather Review*, vol. 135, pp.3506-3520, 2007.

[6] E. Aldrian and R. D. Susanto, "Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature," *International Journal of Climatology*, vol. 23, pp. 1435-1452, 2003.

[7] Buono, I. S. Sitanggang, Mushthofa, and A. Kustiyo, "Time-delay cascading neural network architecture for modelling time-dependent predictor in onset prediction," *Journal of Computer Science*, vol. 10, no. 6, pp. 976-984, 2014.

[8] Indonesian Bureau of Meteorology, Climatology and Geophysics (2013). Prakiraan Musim Hujan [online]. Available:http://www.bmkg.go.id/BMKG_Pusat/DataDokumen/PMH_1314.pdf

[9] E. R. Walpole, "Pengantar Statistika," Jakarta: Gramedia, 1992.

[10] R. Kohavi, "A study of cross validation and bootstrap for accuracy estimation and model selection," *Proceedings of the International Joint Conference on Artificial Intelligence*, vol. 2, pp.1137-1145, 1995.

ISBN : 978-979-1421-225