Comparison of Adaline and Multiple Linear Regression Methods for Rainfall Forecasting

IP Sutawinaya¹, INGA Astawa¹, NKD Hariyanti²

¹Department of Electrical Engineering, Politeknik Negeri Bali, Indonesia
²Department of Business Administration, Politeknik Negeri Bali, Indonesia

E-mail: putusutawinaya@yahoo.co.id

Abstract. Heavy rainfall can cause disaster, therefore need a forecast to predict rainfall intensity. Main factor that cause flooding is there is a high rainfall intensity and it makes the river become overcapacity. This will cause flooding around the area. Rainfall factor is a dynamic factor, so rainfall is very interesting to be studied. In order to support the rainfall forecasting, there are methods that can be used from Artificial Intelligence (AI) to statistic. In this research, we used Adaline for AI method and Regression for statistic method. The more accurate forecast result shows the method that used is good for forecasting the rainfall. Through those methods, we expected which is the best method for rainfall forecasting here.

1. Introduce
The rain intensity is the amount of rainfall per unit within a certain time period. The intensity is said to be great if the rain is heavy and this is very dangerous condition because the impact can cause flood, landslide and negative effects on plants [1]. High rainfall intensity can cause disasters, therefore it is necessary to forecast to estimate how much rainfall will appear. In flood disaster, the main cause of flood is the high intensity of rainfall, so the river capacity is unable to overcome the runoff. Given the rainfall factor is a dynamic factor as a factor causing floods compared with other factors, such as the condition of watersheds and drainage channels, the rainfall is very interesting to continue to study [2]. Forecasting is a process to estimate how many future needs will encompass the needs in quantity, quality, time and location size required to meet the demand for goods or services [3]. Rainfall forecasting can be done with Artificial Intelligence (AI) and statistics [4]. In artificial intelligence, the method used is Adaline, while forecasting with statistics using Regression method. Some research has been done with artificial neural network method [5], [6]. Some research on the application of artificial intelligence or statistics has been done [4]. Method for rainfall prediction was compared to the statistic method. The results show the performance very good which statistic approach and artificial intelligence are quite balance in performance. In [4] Autoregressive Integrated Moving Average (ARIMA) has the worst performance due to its weakness in dealing with non-stationary data. Splines Adaptive Threshold Autoregressive (ASTAR) has the best performance.

The results of the Adaline and Regression methods will be compared by finding the result of the error rate in predicting rainfall. The result of forecasting with a smaller error rate indicates that the...
method is good for forecasting. Expected through comparison of forecasting done with Adaline method which included in Artificial Intelligence (AI) and Regression method which included in Statistic field can know which is better in forecasting.

2. Research Method

2.1. Research Data

Data that used in this research is rainfall data from Kota Denpasar [7]. 2006 to 2016. The data is data per month from each year from 2006 to 2016. The rainfall data that used in this research is shown in Table 1.

Table 1. Rainfall Data in Kota Denpasar from 2006 to 2016

| Year | Rainfall (mm) / Month |
|------|-----------------------|
|      | Jan | Feb | Mar | Apr | Mei | Jun | Jul | Aug | Sept | Oct | Nov | Des |
| 2006 | 436.3 | 193.6 | 301.8 | 227.7 | 77 | 14.5 | 7.1 | 9.1 | 1 | 34.7 | 33.7 | 132.5 |
| 2007 | 119.5 | 94.9 | 425.6 | 96.7 | 25 | 19.9 | 5.4 | 17.4 | 0 | 44.2 | 269.4 | 437.5 |
| 2008 | 325.8 | 364.8 | 344.1 | 111.9 | 64.2 | 1.3 | 1 | 0.7 | 73.2 | 144 | 219.7 | 232.9 |
| 2009 | 561.8 | 412.7 | 261.5 | 22.6 | 73.1 | 2.9 | 9.9 | 0 | 62.6 | 5.3 | 174 | 246.6 |
| 2010 | 304.7 | 307.6 | 36.1 | 359.1 | 228.1 | 124.4 | 120 | 103.1 | 241.6 | 205.6 | 340.6 | 441.2 |
| 2011 | 412.4 | 290.3 | 245.5 | 303.7 | 141.1 | 8.9 | 28.8 | 0.2 | 1.9 | 72.2 | 276.6 | 393.4 |
| 2012 | 730.5 | 168.1 | 554.8 | 18.5 | 77 | 0.2 | 53.2 | 0.2 | 10.9 | 3.8 | 69.6 | 339.4 |
| 2013 | 517 | 144 | 136 | 55 | 143 | 168 | 99 | 0 | 15 | 17 | 234 | 222 |
| 2014 | 360 | 340.5 | 97 | 166 | 68.1 | 65 | 34.3 | 2 | 0 | 1 | 89 | 409 |
| 2015 | 392 | 245.9 | 272.2 | 33.2 | 53 | 0 | 0.5 | 5.8 | 0.7 | 0 | 13 | 157.7 |
| 2016 | 109.4 | 448.1 | 9.7 | 31.1 | 21.4 | 117.2 | 191.4 | 39 | 235.9 | 127.4 | 322.9 | 398 |

The data in Table 1 were normalized before entering the Adaline process as well as the Regression process. Normalization is the process of scaling the attribute value of the data so it can fall in a certain range. The normalization method used in this research is Min-Max method. The Min-Max method used will transform the data into intervals 0 through 1 [8]. Formula 1 is a Min-Max method formula with an interval transformation [0-1].

\[
X' = \frac{0.8(X - b)}{(a - b)} + 0.1
\]

Where \(X'\) is normalized data, \(X\) is real data, \(a\) is maximum value from real data, \(b\) is minimum value from real data

2.2. Adaline

Adaline includes a single-layer network that is part of a feed forward network, where signals come from inputs flowing into output. Adaline is adaptively trained to reduce the error rate [9], [10]. Single layer networks have only one connection layer. In a single layer system, the unit can be clearly distinguished as the unit of input and output unit. Usually in a single layer model each input unit is connected to an output unit but not connected to other input units [9]. Figure 1 Shows the adaline network architecture that has been implemented in this study. In the input layer, it consists of 5 neurons ie data of 2011 (x1), year 2012 (x2), year 2013 (x3), year 2014 (x4), and 2015 (x5). Variable \(W1, W2, \ldots, W51\) is the weight of the output layer and \(W01\) is biased at the output layer. The output layer is denoted by \(y\). Adaline process flow chart can be seen in Figure 2.
Figure 1. Adaline flowchart

Normalized rainfall data and randomly assigned weights and biases that further prepare the maximum limit of iterations. Adaline process begins by calculating the output response unit with the formula:

$$net = \sum_{i=1} x_i w_i + b$$  \hspace{1cm} (2)

Activation function is generated using the formula:

$$Y = f(net) = net$$  \hspace{1cm} (3)

After generating the activation function, the error value starts to be calculated by using:

$$E = (t_i - y_i)$$  \hspace{1cm} (4)

New weight is calculated by the formula:

$$Weight = \alpha (t - y); \quad Bias = \alpha (t - y)$$  \hspace{1cm} (5)

The process of calculating the unit's output response to the change of weight is repeated until it reaches the maximum limit of iteration. Weight and last bias saved.

2.3. Regression

Regression analysis is a mathematical model used to pattern the relationship between two or more variables [11]. Regression method used in this research is multiple regression. Multiple regression can be defined as the intermediate effect of one dependent variable and two or more independent variables. These variables can be used to construct equations and use these equations to make estimates. Multiple regression models modeled response changes. The functional relationship of response y and X with the regression model [11]. Multiple regression models are shown in the formula:
\[ y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (6) \]

Where \( n \) is Number of data, \( y \) is response (dependent variable), \( X_i \) is (independent variable), and \( \beta_i \) is Regression coefficients.

The multiple regression model that was implemented in this study was adjusted into matrix form as the following example:

\[
\begin{bmatrix}
    y_1 \\
    y_2 \\
    \vdots \\
    y_n \\
\end{bmatrix}
= 
\begin{bmatrix}
    1 & x_{11} & x_{12} & \cdots & x_{1n} \\
    1 & x_{21} & x_{22} & \cdots & x_{2n} \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    1 & x_{n1} & x_{n2} & \cdots & x_{nn} \\
\end{bmatrix}
\begin{bmatrix}
    \beta_0 \\
    \beta_1 \\
    \vdots \\
    \beta_n \\
\end{bmatrix}
\quad \text{and} \quad 
\begin{bmatrix}
    \epsilon_0 \\
    \epsilon_1 \\
    \vdots \\
    \epsilon_n \\
\end{bmatrix}
\]

With the formula to calculate the value of regression coefficient as follows:

\[
\hat{\beta} = (X'X)^{-1}X'y \quad (7)
\]

Where \( X' \) is transpose form of \( X \), and \((-1)\) is inverse. So the model can be represented as:

\[
\hat{y} = X\hat{\beta} \quad (8)
\]

and error value \( e \) obtained with,

\[
e = y - \hat{y} \quad (9)
\]

The flow diagram for the multiple regression process is shown in Figure 2. The flow diagram begins with the normalization of rainfall data. The normalized data are then searched for regression coefficient values. To find the regression coefficients, it is necessary to convert the equations to the matrix equations as in the formulas 7 through 9. After the coefficients are obtained, the error values can be searched.

3. Discussion

3.1. Normalizing data

The data that contained in Table 1 is normalized according to formula 1. The range of values in normalization is \([0,1]\). The learning phase (training) uses normalized rainfall inputs from 2006 to 2014. The target is rainfall data in 2015, both on Adaline and Multiple Linear Regression methods using the data. The input data used at the time of testing is the data in 2007 to 2015 and which is predicted to be 2016. Benchmarking of Adaline and Multiple Linear regression methods is seen from the difference in predictions from each method to the actual data.
3.2. Adaline
Adaline method that conducted in this research is using learning rate of 0.05 and maximum iteration as much as 1000. Testing that has been done as previously mentioned has produces data. The graph for predicted results is shown in Figure 3.

3.3. Multiple linear regression
Forecasting test was conducted using multiple linear regression method get results. In this test, in multiple linear regression methods there are 9 independent variables has used. The independent
variables $x_1$ to $x_9$ are rainfall data from 2007 to 2015. The graph for predicted results is shown in Figure 4.

![Figure 4. Comparison graph of actual rainfall data and prediction using Multiple linear regression](image)

3.4. Comparison

The summary of the predicted results along with the MSE and RMSE values shown in Table 2.

**Table 2. Summary of rainfall forecasting results in 2016 along with MSE and RMSE values**

| month | $y$   | $y'$  | $y-y'$ | $(y-y')^2$ | $y$   | $y'$  | $y-y'$ | $(y-y')^2$ |
|-------|-------|-------|-------|------------|-------|-------|-------|------------|
| 1     | 0.219808 | 0.361422 | -0.14161 | 0.020054 | 0.219808 | 0.364887 | -0.14508 | 0.021048 |
| 2     | 0.590732 | 0.251518 | -0.339215 | 0.115067 | 0.590732 | 0.238302 | -0.35243 | 0.12407 |
| 3     | 0.110623 | 0.249457 | -0.13883 | 0.019275 | 0.110623 | 0.256485 | -0.14586 | 0.02176 |
| 4     | 0.134059 | 0.17669  | -0.04263 | 0.001817 | 0.134059 | 0.177698 | -0.04364 | 0.001904 |
| 5     | 0.123436 | 0.206919 | -0.08348 | 0.006969 | 0.123436 | 0.215644 | -0.09221 | 0.008502 |
| 6     | 0.22835  | 0.193542 | 0.034808 | 0.001212 | 0.22835  | 0.203346 | 0.025004 | 0.000625 |
| 7     | 0.30961  | 0.172523 | 0.137086 | 0.018793 | 0.30961  | 0.185467 | 0.124143 | 0.015412 |
| 8     | 0.14271  | 0.158436 | -0.01573 | 0.000247 | 0.14271  | 0.17094  | -0.02823 | 0.000797 |
| 9     | 0.358344 | 0.199238 | 0.159106 | 0.025315 | 0.358344 | 0.214339 | 0.144004 | 0.020737 |
| 10    | 0.239521 | 0.167978 | 0.071543 | 0.005118 | 0.239521 | 0.179616 | 0.059905 | 0.003589 |
| 11    | 0.453621 | 0.269568 | 0.184053 | 0.033875 | 0.453621 | 0.263107 | 0.190514 | 0.036296 |
| 12    | 0.535866 | 0.303899 | 0.231966 | 0.053808 | 0.535866 | 0.297041 | 0.238825 | 0.057037 |

RMSE = 0.158522  RMSE = 0.161098
MSE = 0.025129   MSE = 0.025953

In Table 2, the MSE values obtained from the rainfall forecasting test in 2016 using Adaline and Multiple Linear Regression method are 0.025129 and 0.025953 respectively. The RMSE values obtained from the Adaline method and Multiple linear regression are 0.158522 and 0.161098 respectively.
4. Conclusion
Comparison of rainfall forecasting test using Adaline method with Multiple linear regression method which has been done in this research is not too different in error rate. By using rainfall data of Kota Denpasar from year 2006 until 2016, where data from year 2006 until 2015 become base of learning in each method. Data from 2007 to 2016 became the test data in this study. The error rate in the form of MSE value obtained from rainfall forecasting test in 2016 using Adaline and Multiple Linear Regression in sequence of 0.025129 and 0.025953. The RMSE values obtained from the Adaline method and multiple linear regression are 0.158522 and 0.161098 respectively. Adaline methods that included in the field of Artificial Intelligence (AI) get a smaller error rate value. That is, the Adaline method is better used for rainfall forecasting than Multiple linear regression methods because of the smaller error rate.

5. Acknowledgments
This research was support by grants of DIPA Politeknik Negeri Bali Sp Dipa-042.01.2.401006/2017.

6. References
[1] Jun-He Y, Ching-Hsue C, 2015. A novel rainfall forecast model based on integrated non-linear attributes selection method and support vector regression. 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD).
[2] Konrad CP. 2016 Effects of Urban Development on Floods. US GEOLOGICAL SURVEY.
[3] Moon MA. 2015. Demand and Supply Integration: The Key to World-Class Demand Forecasting New Jersey 07458: FT Press.
[4] Amirullah I, Harun N, Pallu MS, Arwan A. 2013 Statistic Approach versus Artificial Intelligence for Rainfall Prediction Based on Data Series. International Journal of Engineering and Technology (IJET).5(2):1962-9.
[5] Elsafi SH. 2014 Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan. Alexandria Engineering Journal.53(3):655-62.
[6] Gholizadeh MH, Darand M. 2009 Forecasting Precipitation with Artificial Neural Networks (Case Study: Tehran). Journal of Applied Science.9(9):1786-90.
[7] BMKG. 2015 Data Online Pusat Database BMKG. Available from: http://dataonline.bmkg.go.id/data_iklim.
[8] Han J, Kamber M, Pei J. 2012. Data Mining Concepts and Techniques 3rd Edition Elsevier.
[9] Pellakuri V, Rao DR, Murthy JVR, 2016. Modeling of supervised ADALINE neural network learning technique. 2nd International Conference on Contemporary Computing and Informatics (ICCI).
[10] Yang W. 2000 A comparison of training techniques : ADALINE, back propagation and genetic algorithms: Concordia University
[11] Li H, Pi D, Wu Y, Chen C. 2017 Integrative Method Based on Linear Regression for the Prediction of Zinc-Binding Sites in Proteins. IEEE Access.5:14647-56.