Implementation of artificial neural networks for very short range weather prediction

G B Wanugroho¹, Martarizal* and R M Putra²

¹Departement of Physics, Universitas Indonesia, Depok, Indonesia
²Indonesian Agency for Meteorology Climatology and Geophysics

*martarizal@sci.ui.ac.id

Abstract. Weather conditions are a significant factor for various sectors such as transportation safety, development, health, etc. Therefore, high development is needed in forecasting future weather conditions. Many ways are used to predict weather conditions. Along with the development of technology now, weather prediction can be made using Artificial Intelligence (AI) technology or artificial intelligence so that the results obtained are more optimal. In this study, the artificial neural network used has a feedforward neural network algorithm using training data consisting of temperature, air pressure, air humidity, wind speed, hourly wind speed at the Juanda Meteorological Station in Surabaya in January 2019 with the target is intensity of rainfall. Furthermore, the data was released in the period of 1 January 2019 to 31 January 2019. Based on the analysis results, the Artificial Neural Network model has a fairly good performance in predicting an increase in rainfall in Surabaya. The best model is considered by a model with architecture 7 - 60 - 1 with an estimate correlation is 0.87, with an error value of -0.03. With this model, it is expected to become one of the forecaster considerations in making special weather forecasts at intervals every hour.

1. Introduction

Weather information services are now needed by almost all walks of life for daily use. BMKG (Climatology Meteorology and Geophysics Agency) is the only government agency in Indonesia that has a duty to provide weather information to the wider community, the information disseminated must be accurate and can be accounted for both in terms of accuracy and legality. As one of the countries in the tropics, Indonesia has the potential to experience extreme weather which is a temporary (momentary) event at various scales of disturbance (Handayani, 2010) [1]. According to BMKG, extreme weather is an unusual and unusual weather event that can result in losses, especially lives and property. Rain is classified in the extreme weather category if it has the lowest intensity of 50mm / day and or 20 mm / hour (BMKG, 2010) [2].

The territory of Indonesia, which among others by the ocean + - 2/3 parts, as well as high heating originating from heating sunlight, rich in water vapor and high humidity, physically becomes fertile ground for cloud convective formation that causes heavy rainfall (Kurniawan et.al, 2004) [3]. One factor that can generally describe the weather conditions is rainfall. But rainfall is a weather factor that is difficult to predict accurately. Rainfall that occurs can not be determined precisely, but can be predicted or estimated. By using rainfall data in the past it can be used to predict the rainfall that will occur.
Artificial Neural Networks (ANN) is one method that can be used to estimate weather conditions. ANN is able to make an introduction to past data-based activities. Data will be studied by artificial neural networks so that it has the ability to make decisions on data that has never been studied. ANN also provides a methodology for solving various types of non-linear problems (Abraham et al 2004) [4].

Research that examines weather prediction by utilizing Artificial Neural Networks has been carried out. One of them is Handayani and Adri’s research [5] which examines the application of ANN for the prediction of rainfall in the city of Pekanbaru using data on air temperature, humidity, air pressure and wind speed in the period 2012-2013. This study uses a backpropagation method using the activation of binary sigmoid and bipolar sigmoid functions. The results showed that the introduction of data patterns with backpropagation ANN has a small error value. In general, this study concludes that parameters such as learning levels, goals, activation functions and hidden layers greatly influence the outcome of rainfall predictions. Similar research was conducted by Dewi, Bhari and Irwansyah [6] using ANN method with backpropagation algorithm consisting of 1 input layer with 7 neurons, 1 hidden layer with 12 neurons, and 1 output layer with 1 neuron. This study, which uses daily data on air temperature, humidity, pressure and rainfall at the Lombok International Airport Meteorological Station for the period January 2015 - December 2016, shows the results of the introduction of a pretty good pattern with UMK 25.0639.

In this study, the scope is limited, including: weather parameters predicted using parameters of temperature, dew point, pressure, humidity, wind speed, and rainfall, the research location is Juanda Surabaya Meteorological Station, the data used are observational data (synop) of January 1, 2019 to January 31, 2019. The algorithm used is a feedforward neural network consisting of 10 models of Artificial Neural Networks as in table 1.

2. Methodology

2.1. Research Data
This study uses hourly weather observational data from the Surabaya Juanda Meteorological Station which was obtained directly from the Juanda Meteorological Station Surabaya trainer. The training data period used for this research material is 1 month, from 1 January to 31 January 2019. The initial step is to do quality control by looking at all the availability and quality of data. If there is one element that is missing or invalid data, then that day's data will be deleted. Furthermore, the model that has been created with the training data will be tested in the case of January 2019 to see how the performance of the model. The Meteorology Climatology and Geophysics Agency divides the rainfall intensity criteria as follows:

| No | Amount of Rain (mm) | Rain Criteria   |
|----|---------------------|-----------------|
| 1. | 1.0 – 5.0 mm/hour or 5.0 – 10.0 mm/day | Slight Rain     |
| 2. | 5.0 – 10.0 mm/hour or 20.0 – 50.0 mm/day | Moderate Rain   |
| 3. | 10.0 – 20.0 mm/hour or 50.0 – 100.0 mm/day | Heavy Rain      |
| 4. | > 20.0 mm/hour or >100 mm/day | Very Heavy Rain |

(Source: BMKG)

2.2. Artificial Neural Network
At this stage is the design of the system used as training data in each forecast chosen at the research boundary. The first case study location is determined which can be used as a research sample so that the weather observation data to be used is more optimal. Furthermore, weather observation data from UPT BMKG Juanda Surabaya with parameters of temperature, dew point, air pressure, humidity, wind speed and rainfall are collected and arranged in a format that can be detected by the matlab
application. In conducting training and testing the results of weather data processing, the data processing will also be tested into a computerized system. The software will be used in testing the data to determine the weather prediction process the next day by using Matlab software.

Before conducting model training, the architectural design of the model must be determined. In this study, there are 10 design models to be tested. In addition, variations in the number of neurons in each hidden layer were made. This was done to test the effect of the number of hidden layers and neurons on the output quality results of the model created. Table 1 is a detailed description of the description of the ANN model architecture to be made.

Table 2. ANN Model Architectural Design

| Model | Number of Hidden Layer | Model Architectural Design | Model | Number of Hidden Layer | Model Architectural Design |
|-------|------------------------|---------------------------|-------|------------------------|---------------------------|
| I     | 1                      | 7 – 30 – 1                | VI    | 2                      | 7 – 40 - 40 – 1           |
| II    | 1                      | 7 – 40 – 1                | VII   | 2                      | 7 – 50 - 50 – 1           |
| III   | 1                      | 7 – 50 – 1                | VIII  | 2                      | 7 – 60 - 60 – 1           |
| IV    | 1                      | 7 – 60 – 1                | IX    | 1                      | 7 – 100 – 1              |
| V     | 2                      | 7 – 30 - 30 – 1           | X     | 2                      | 7 – 100 - 700 – 1        |

The steps that must be taken for the Feedforward Neural Network method with matlab is to make a network initiation. The data will be used using matlab with architectural patterns as in table 1, with input of seven weather parameters used, 10 kinds of hidden layers and a total of 1 output. The aim is to create and train a network that can predict components of weather parameters both from the previous day (target data) and historical data one year ago (input data).

Figure 1. Research Flow

The error rate of the model can be determined by calculating the Mean Absolute Error value from the results of the model [7]. The greater the MAE value, the greater the error rate of the output, indicating that

\[ MAE = \frac{\sum_{i=1}^{n}|RM_i - RO_i|}{n} \]  

Information:

MAE = Mean Absolute Error  
RM = Rainfall Model  
RO = Rainfall Observation

The model is less than optimal in making the rain forecast. Mathematically, the value of an MAE can be calculated as follows:

In addition, seen how the relationship with the output model that has been made with the results of observation. The level of the relationship can be calculated based on the correlation value [7].

\[ C = \frac{\sum_{i=1}^{n}(RO_i - \bar{RO})(RM_i - \bar{RM})}{\sqrt{\sum_{i=1}^{n}(RO_i - \bar{RO})^2} \sqrt{\sum_{i=1}^{n}(RM_i - \bar{RM})^2}} \]  

Information:

RM = Rainfall Model
$RM = $ Average of Model Rainfall Intensity

$RO = $ Observation intensity

$RO = $ Average of Observation Rainfall

The greater the value of C (correlation), the relationship between the two data being compared is getting stronger. The last test was performed with the accuracy value of the rainy day event output (rain $\geq$1 mm). The accuracy value shows the performance of the model to provide an forecast of rainfall conditions for 1 day. To calculate the accuracy value can be done with the following equation:

$$Accuracy = \frac{H}{N}$$

Information:

$H = $ results of the model and observation have the same conditions (rain and rain; no rain and no rain)

$N = $ Overall total data

3. Result and Discussion

Based on the results of hourly weather data training on weather conditions in the form of rainfall intensity, ANN model has a fairly good performance in detecting rainfall events. In the whole model, the pattern of events of rainfall intensity during the January 2019 period has similar characteristics, but there are some conditions which are not suitable for the output of the ANN model that has been made.

Furthermore, the correlation value of the outputs of all models with the result of observational rainfall ranges between 0.55 - 0.87. This shows that the level of correlation between the output model and the results of observations is being to strong [7]. The lowest correlation value is found in model II, which is a model with 1 hidden layer with 40 neurons in that layer. While the highest correlation is found in Model IV which has 1 layer of neurons with a total of 60 neurons.

![Figure 2. Results of comparison of outputs for all models and the best model](image)

![Figure 3. The Best Model (ANN4)](image)
Based on the whole previous analysis, the most optimal model in this case study is shown by Model IV, which is a model with 1 hidden layer that has 60 neurons. In addition, the increasing number of neurons and the number of hidden layers is not always directly proportional to the output results that are getting closer to the results of observation. Based on the entire discussion, model IV has the smallest error rate of -0.03, the highest correlation is 0.87, and the accuracy rate of rainy days is 95%.

Furthermore, seen from the error value generated by the model compared with the results of observation. The error rate of the model is indicated by the Mean Absolute Error value. If the MAE value is close to 0, then the results indicate that the model has good performance. In Figure 4, is the MAE value based on all output models that have been made. These results indicate that in general, the error value in all models is below 10 mm, with an error range between -1.65 mm - 1.0 mm. MAE values have a non-uniform pattern when the number of hidden layers and the number of neurons are added. The smallest error rate is shown in Model IV which has 1 hidden layer with 60 neurons. While the biggest error is shown in Model VII which has 2 hidden layers with a total of 50 neurons in the first layer and the second layer.

Subsequent tests were carried out to see the accuracy of the model in predicting rainfall intensity conditions based on the new input data provided. Based on the outputs of Model I through Model X, the accuracy of rain conditions with a value of ≥1 mm has a value that varies with a range of 78% - 96%. For the model with the highest level of rainy day accuracy shown by Model III, which consists of 1 hidden array with 50 neurons. Then the next best accuracy is shown by Model IV with an accuracy value of 95%. The lowest accuracy value is shown by Model VII with an accuracy value of 78% which has 2 hidden layers with a total of 50 neurons each.

**Figure 4. Correlation Model Value and Model Error Rate**

**Figure 5. Accuracy of Predicted Rainfall Conditions Models**
4. Conclusion
Based on the results of research conducted using training data during January 2019, the results of the model using the Artificial Neural Network are quite good and close to the observation rain conditions. In this study, increasing the number of neurons in the hidden layer and the number of hidden layers has a significant influence on the output of the model. However, there is no specific pattern that indicates that the number of neurons and the number of hidden layers increases. The best model results in this study are shown by ANN Model with 7-60-1 architecture showing the smallest error level is 0.03 mm and the highest correlation is 0.87. In addition, this mode also has a high accuracy of rainy day events, namely 96% in the test data period, namely January 2019.

References:
[1] Handayani A. S. 2010. Analisis daerah endemic bencana akibat cuaca ekstrim di Sumatera Utara.
[2] BMKG, 2010. Keputusan no. 009 Tentang Prosedur Standar Operasional Pelaksanaan Peringatan Dini, Pelaporan, dan Diseminasi Informasi Cuaca Ekstrem. BMKG, Jakarta.
[3] Kurniawan, R. 2012. Kajian Daerah Rawan Gelombang Tinggi Di Perairan Indonesia (Studi On High Wave Prone Areas Over Indonesian Waters)
[4] Abraham A. Maqsoed I. Kahn m.2004. An Ensemble of neural networks for weather forecasting. Neural Comput % Apllic (2004) 13: 112-122
[5] Handayani L dan Adri M. 2015. Penerapan JST (backpropagation) untuk Prediksi Curah Hujan (Studi Kasus : Kota Pekanbaru). Prosiding Seminar Nasional Teknologi Informasi, Komunikasi, dan Industri (SNTIKI) 7.
[6] Dewi K.N.A., Bhari S., dan Irwansyah. 2019. Model Prediksi Curah Hujan Harian Menggunakan Jaringan Syaraf Tiruan Backpropagation. Indonesia Physical Review. 2(1).
[7] Ali A. dan Putra R.M. 2014. Identifikasi Kejadian Hujan Es Menggunakan Radar Cuaca Doppler (Studi Kasus Hujan Es Jakarta 22 April 2014 dan Denpasar 16 Desember 2010). Prosiding SNSA 2015. Hal 1 - 6