

Rainfall Prediction Using Backpropagation with Parameter Tuning

Wahyudi Setiawan^{1*}, Asyroful Barokah², and Mula'ab³

^{1,2,3}Department of Informatics, University of Trujonoyo Madura

Abstract. Rainfall is one of the important elements in the process of weather and climate. The high intensity of rainfall every year can hamper the mobility of the population and the distribution of goods, especially in the port area. Rainfall prediction is needed to handle the impacts caused by high rainfall. The data was obtained from the website dataonline.bmkg.go.id with observations made by the Tanjung Perak Surabaya Maritime Meteorological Station. The prediction method uses an artificial neural network with Backpropagation. Autocorrelation function is used to determine the number of input neurons with the best features in the Artificial Neural Network. Rainfall data is divided into two parts,; January 2008 to December 2019 used for training data and January to August 2020 for testing data. The validation technique used is 10-Fold Cross Validation. The experiment uses parameter tuning of iteration and learning rate. The training process obtained the best learning rate was 0.2 and 1000 iterations with a MSE validation score of 0.02591. Finally, the testing process has a Mean Square Error value of 0.02769 and a percentage of true rain character of 62.5%.

Keywords. Autocorrelation Function, Backpropagation, Cross Validation, Rainfall prediction.

1 Introduction

Rainfall is one of the important elements in the process of weather and climate. Rainfall is the amount of rainwater that accumulates on a flat surface and does not evaporate, seep out, or flow. Rainfall is one millimetre, meaning that in an area of one square meter on a flat place one millimetre of water can be accommodated [1]. Monthly rainfall information is needed in irrigation or water construction planning, infrastructure planning, agriculture, transportation, and also telecommunications. Monthly rainfall has the character of rain which is a comparison between the amount of rain that occurs during one month with the average or normal value of that month. Rainfall characters are divided into three categories above-normal, normal, or below-normal [2], [3].

High monthly rainfall certainly has a negative impact on an area. One of the impacts is the inhibition of population mobility and the distribution of goods so that economic development in other areas is also disrupted [4], [5]. One of the facilities for population mobility and economic support in an area is a port. The high intensity of monthly rainfall results in delays in the up and down of passengers and the process of loading and unloading goods. So we need prediction information to calculate the intensity of monthly rainfall and the character of the rain. These predictions are used to prepare for handling to minimize losses caused by high rainfall in a month.

Rainfall that occurs in an area cannot be determined with certainty, but its intensity can be predicted using historical rainfall data.

Research to predict rainfall has been carried out with various methods and certain applications. In a previous study on the prediction of monthly rainfall with the Backpropagation Neural Network method, the MSE was 5.590×10^{-32} [3]. Then another study also used the Backpropagation Artificial Neural Network, the network succeeded in recognizing training patterns of 99.0% [6]. Both studies used 10 years of data for training and 1 year for testing. As well as using 10 inputs based on 10 years of data used with the same month data in previous years.

The studies that have been done show that the Backpropagation method gives good results for prediction. The monthly rainfall prediction in this study uses monthly rainfall data from January 2008 to August 2020 in the port area of Tanjung Perak Surabaya which was observed by the Tanjung Perak Surabaya Maritime Meteorology Station. This study uses the Artificial Neural Network to predict monthly rainfall. Backpropagation to improve the weights so that the next forward propagation can produce a better predictive value. It is also used parameter tuning for learning rate and iteration. Before being processed with the Neural Network, data analysis is applied using the Autocorrelation Function to determine the input layer neurons.

* Corresponding author: wsetiawan@trunojoyo.ac.id

2 Research Methods

2.1 Autocorrelation Function

The Autocorrelation function is a linear relationship between observations Z_t with Z_{t-k} [7], [8]. The Autocorrelation function calculates the correlation between variables from a time series dataset that is distinguished from its time unit. The autocorrelation function is chosen to be the strongest correlation to the input neuron in the Backpropagation. The autocorrelation function formula is found in equation 1 [9].

$$r_k = \frac{\sum_{i=1}^{n-k} (Z_t - \bar{Z})(Z_{t-k} - \bar{Z})}{\sum_{i=1}^n (Z_t - \bar{Z})^2} \quad (1)$$

r_k = lag times; Z_t = data now; Z_t = average data; Z_{t-k} = data k (-1 and so on)

2.2 Rainfall

Rainfall is the amount of water that falls from rain in a particular time and area. Usually expressed as a virtual depth of coverage. [1]. Rainfall data at certain locations can be accessed on the Database Center Online Data page – BMKG dataonline.bmkg.go.id. Rainfall criteria in the territory of Indonesia [10]:

- 0 mm/day: Cloudy
- 0.5 – 20 mm/day: Light rain
- 20 – 50 mm/day: Moderate rain
- 50 – 100 mm/day: Heavy rain
- 100 – 150 mm/day: Very heavy rain
- >150 mm / day: Extreme rain.

2.3 Character of rain

The character of rain is the ratio between the amount of rain that occurs during one month with the average or normal value of that month in a place. There are three criteria [6]:

- Above-Normal (AN), if the comparison value is > 115% of the average rain.
- Normal (N), if the comparison value is between 85% - 115% of the average rain.
- Below-Normal (BN), if the comparison value is < 85% of the average rain.

The formula for the character of rain is shown in equation 2

$$\text{character of rain month } n = \frac{\text{total rainfall month } n}{\text{average rainfall month } n} \times 100\% \quad (2)$$

2.4 Backpropagation

Backpropagation is a multi-layer that changes the weight value by going backward from the output layer to the input layer. By changing the weight value during the training process, A backpropagation architecture is

an architecture that can be used to examine and analyse historical data patterns in more detail and obtain output (with minimal error). The Backpropagation training algorithm consists of three phases [11], [12]:

- Input the value of the training data so that the output value is obtained.
- Backpropagation of the obtained error value.
- Adjusted the correction weight to minimize the error value.

2.5 Means Square Error (MSE)

The results issued in the output layer are compared with the target values and calculated using the Mean Square Error (MSE). Each error or remainder is squared. In the training process there is a stop condition if the MSE value \leq the minimum error limit [13].

2.6 Dataset

The Dataset used in this study is data on the amount of monthly rainfall observed from the BMKG Tanjung Perak Surabaya Maritime Meteorological Station which was obtained from the website <http://dataonline.bmkg.go.id/> from January 2008 to August 2020.

2.7 System Proposed

The design of the system consists of preprocessing for determining inputs, then training to produce the best training weights, and testing used to generate predictions and evaluate the result. The proposed system can be seen in Figure 1. Fig. 1. System Architecture Proposed

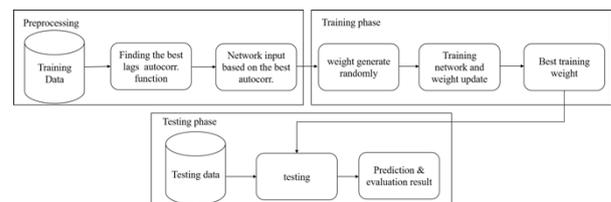


Fig. 1. System Architecture Proposed

In the preprocessing, training data has the monthly rainfall from 2008–2019. The best correlation is sought with the data 1 month to 24 months backward using the Autocorrelation Function. From the autocorrelation results, the five best correlations are between the main data (2008 – 2019) with data that is backward -1, -11, -12, -13, and -24 months. So, the inputs are 1, 11, 12, 13, and 24 from the monthly rainfall target data. There is input data that is 24 months backward from the target, the total training data is rainfall data. January 2008 to December 2019.

In the training process, the data is used from the results of the Autocorrelation Function. Further generated random weights, a forward propagation process is carried out. Calculating the error between the output predicted and the output target. A backward propagation process is carried out to update the weights. After the

training process is carried out, the resulting weights (best training weight) are used for the testing process. The testing data used is rainfall data from January to August 2020 based on the Autocorrelation Function. The forward propagation is carried out using the weights from the training results. The testing produces rainfall predictions and calculates the evaluation of the prediction using the Mean Square Error (MSE) and the percentage of true monthly rainfall character

3 Result and Discussion

3.1 Training and Testing Parameter

Parameters in the training process include using three layers: the input layer, hidden layer, and output layer. There are five input neurons based on the results of the Autocorrelation Function analysis. Also, four hidden neurons are based on 70-90% of the number of input neurons. The output neurons are based on the predicted monthly rainfall. The learning rate parameter tuning is from 0.2 until 0.8. It is to find the best learning rate. The error limit is 0.001, and the maximum number of iterations also uses parameter tuning from 10 to 1000 iterations. It is the aim of finding the best iteration. Activation using sigmoid. The weights used in the testing process are the stored weights from the best training results. The experiment Scenario was carried out using 10-Fold Cross Validation, the number of learning rates, and the number of iterations. At this phase, the system is tested using 120 data lines (2008-2019). There are 10 folds with details of 108 rows of data as training and 12 rows of data as validation.

3.2 Result of Training and Testing

Figure 2 shows the results of the training carried out using the 10-Fold Cross Validation with parameters of three layers, five input neurons, four hidden neurons, one output neuron, learning rate 0.2 until 0.8, 1000 iterations, error limit 0.001, and using sigmoid activation. From the training process, the best results obtained MSE 0.02591.

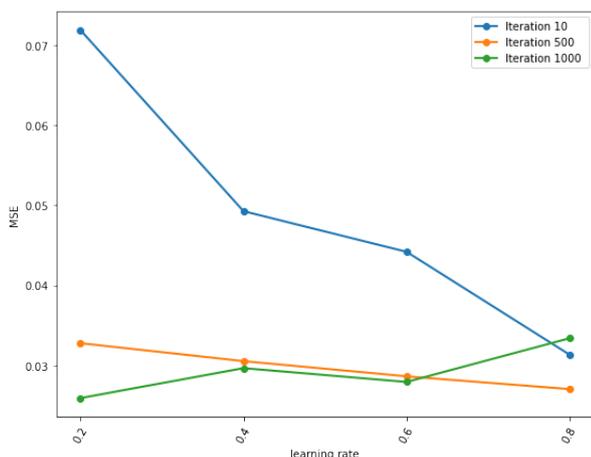


Fig. 2. Mean Square Error Result

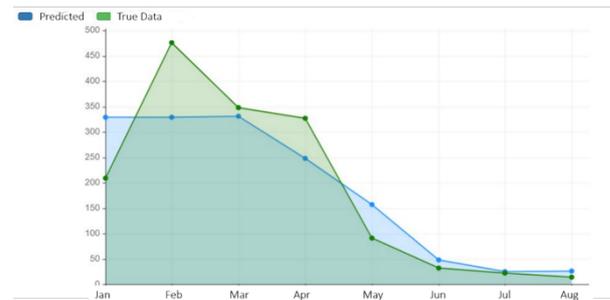


Fig. 3. Actual and predicted data results

Figure 3 shows the results of testing data carried out with the best-stored model. The graph consists of two: green is the actual data and blue is the predicted data. Table 1 shows the comparison of actual and predicted data in 2020. It also shows MSE for each month. Hereafter in Table 2 shows the character of the actual and predicted data.

Table 1. Result of Rain Actual and Predicted Character

Month	Actual Data	Predicted Data	MSE
January	209	329	0.0676
February	476	329	0.10049
March	348	331	0.00137
April	327	248	0.02958
May	91	157	0.02045
June	32	48	0.00122
July	22	25	0.00005
August	14	26	0.00073
MSE			0.02769

Table 2. Result of Rain Actual and Predicted Character

Month	Rainfall avg.	Actual rain	Cat. actual	Predicted rain	Cat. predicted	True /False
Jan	315.8	66.2	BN	104.2	N	F
Feb	309.2	153.9	AN	106.4	N	F
Mar	264.7	131.5	AN	125	AN	T
Apr	187.5	174.5	AN	132.3	AN	T
May	103.9	87.6	N	151.1	AN	F
Jun	57.4	55.7	BN	83.6	BN	T
Jul	23.3	94.4	N	107.3	N	T
Aug	12.1	115.7	AN	214.9	AN	T

AN: Above-Normal; BN: Below-Normal; T: True, F: False

The design of the system consists of preprocessing for determining inputs, then training to produce the best training weights, and testing used to generate predictions and evaluate the result. The proposed system can be seen in Figure 1 System Architecture Proposed

In the pre-processing, training data has the monthly rainfall from 2008–2019. The best correlation is sought with the data 1 month to 24 months backward using the Autocorrelation Function. From the autocorrelation results, the five best correlations are between the main

data (2008 – 2019) with data that is backward -1, -11, -12, -13, and -24 months. So, the inputs are 1, 11, 12, 13, and 24 from the monthly rainfall target data. There is input data that is 24 months backward from the target, the total training data is rainfall data. January 2008 to December 2019.

In the training process, the data is used from the results of the Autocorrelation Function. Further generated random weights, a forward propagation process is carried out. Calculating the error between the output predicted and the output target. A backward propagation process is carried out to update the weights. After the training process is carried out, the resulting weights (best training weight) are used for the testing process.

The testing data used is rainfall data from January to August 2020 based on the Autocorrelation Function. The forward propagation is carried out using the weights from the training results. The testing produces rainfall predictions and calculates the evaluation of the prediction using the Mean Square Error (MSE) and the percentage of true monthly rainfall character

Table 2 shows the results of the actual and the predicted rainfall. The rainfall character of the Above-Normal, Normal, or Below-Normal categories. The rainfall character is obtained from comparison calculations with the average rain as shown in equation 2. It has five months with the same rainfall character (March, April, June, July, and August). So the percentage of true rainfall is 62.5%.

4 Conclusion

The rainfall prediction system was carried out. The parameters of backpropagation consist of three layers, five neurons input layer, four neurons hidden layer, one neuron output layer, sigmoid activation, and parameter tuning of learning rate & iteration. Network input based on $xt - 1$, $xt - 11$, $xt - 12$, $xt - 13$, $xt - 24$ Autocorrelation function and using 10-Fold Cross Validation. The best training results are in the 5th fold out of 10 folds with a validation MSE of 0.02591. Based on the system testing, it is concluded that the results of monthly rainfall data in 2020 produce a Mean Square Error (MSE) of 0.02769. A comparison of the character of rainfall on the actual and predicted data obtained five true rainfall character. The characteristics of true rainfall include March, April, June, July, and August with an accuracy of 62.5%.

The prediction results on the 2020 rainfall are less precise than the actual data caused by monthly rainfall data which has a value range from zero to hundreds. Based on the research that has been done, suggestions that need to improve the research include the experiment using more than one hidden layer. Also, increase the period time used in the analysis of the Autocorrelation Function for the input neuron of more than five neurons.

References

- [1]. L. Q. Avia, "Change in rainfall per-decades over Java Island, Indonesia," in IOP Conference Series: Earth and Environmental Science, (2019), vol. **374**, no. 1, doi: 10.1088/1755-1315/374/1/012037.
- [2]. F. Lebdi, "Irrigation for agricultural transformation - Background paper for African transformation report 2016: Transforming Africa's agriculture," (2016). [Online]. Available: <https://www.jica.go.jp/jica-ri/publication/booksandreports/175nbg0000004aet-att/175nbg0000004ai9.pdf>.
- [3]. D. M. Midyanti, R. Hidayati, and S. Bahri, "Prediksi Curah Hujan Bulanan meteorologi Supadio Kubu menggunakan Jaringan Syaraf Tiruan Backpropagation," in Seminar Nasional II Penerapan Ilmu Pengetahuan dan Teknologi Universitas Tanjungpura, (2016), no. November, pp. 363–368.
- [4]. Asian Development Bank, "Indonesia Climate Risk Country Profile," 2021. [Online]. Available: www.worldbank.org.
- [5]. H. Abdul Rahman, "Climate Change Scenarios in Malaysia: Engaging the Public," *Int. J. Malay-Nusantara Stud.*, vol. **1**, no. 2, pp. 55–77, (2018), [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1086.6702&rep=rep1&type=pdf>.
- [6]. C. Oktaviani and A. Afdal, "Prediksi Curah Hujan Bulanan menggunakan Jaringan Syaraf Tiruan dengan Beberapa Fungsi Pelatihan Backpropagation (Studi Kasus: Stasiun Meteorologi Tabing Padang, Tahun 2001-2012)," *J. Fis. Unand*, vol. **2**, no. 4, pp. 228–237, (2014), [Online]. Available: <http://jfu.fmipa.unand.ac.id/index.php/jfu/article/view/49>.
- [7]. D. A. Rezaldi and Sugiman, "Peramalan Metode ARIMA Data Saham PT . Telekomunikasi Indonesia," in *Prisma*, (2021), vol. **4**, pp. 611–620.
- [8]. A. Assidiq, P. Hendikawati, and N. Dwidayati, "Perbandingan Metode Weighted Fuzzy Time Series, Seasonal Arima, dan Holt-Winter's Exponential Smoothing untuk Meramalkan Data Musiman," *UNNES J. Math.*, vol. **6**, no. 2, pp. 129–142, (2017), [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ujm>.
- [9]. A. Machmudin and B. S. S. Ulama, "Peramalan Temperatur Udara di Kota Surabaya dengan Menggunakan ARIMA dan Artificial Neural Network," *J. Sains dan Seni ITS*, vol. **1**, no. 1, pp. 118–123, (2012).
- [10]. BMKG, "Probabilistik Curah Hujan 24 Jam," [bmkg.go.id](https://www.bmkg.go.id), (2022). [https://www.bmkg.go.id/cuaca/probabilistik-curah-hujan.bmkg#:~:text=0.5 - 20 mm%2Fhari,\(merah\)%3A Hujan sangat lebat](https://www.bmkg.go.id/cuaca/probabilistik-curah-hujan.bmkg#:~:text=0.5 - 20 mm%2Fhari,(merah)%3A Hujan sangat lebat) (accessed Jul. 17, 2022).
- [11]. V. Nagori, "Fine tuning the parameters of back propagation algorithm for optimum learning performance," in *Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics, IC3I 2016*, (2016), pp. 7–12, doi: 10.1109/IC3I.2016.7917926.
- [12]. D. A. Permatasari and D. A. Maharani, "Backpropagation neural network for tuning PID pantilt face tracking," in *Proceedings - 2018 3rd International Conference on Information Technology, Information Systems and Electrical Engineering*,

ICITISEE 2018, (2018), no. November, pp. 357–361,
doi: 10.1109/ICITISEE.2018.8720968.

- [13]. A. Gusi-Amigo, P. Ciosas, and L. Vandendorpe, “Mean square error performance of sample mean and sample median estimators,” *IEEE Work. Stat. Signal Process. Proc.*, vol. **2016-August**, no. June, (2016), doi: 10.1109/SSP.2016.7551739.