ISSN 2278-3091
Volume 8, No.1.5, 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse4881.52019.pdf
https://doi.org/10.30534/ijatcse/2019/4881.52019

Artificial Neural Network for Rainfall Prediction Base on Historical Rainfall Data by Day

Berlilana[1], Wiga Maulana Baihaqi[2], Sarmini[3]

[1]Department of Information System, Faculty of Computer Science, Universitas Amikom Purwokerto, Purwokerto, Indonesia, berli_amikom@yahoo.co.id
[2]Department of Information Technology, Faculty of Computer Science, Universitas Amikom Purwokerto, Purwokerto, Indonesia, wiga@amikompurwokerto.ac.id
[3]Department of Information System, Faculty of Computer Science, Universitas Amikom Purwokerto, Purwokerto, Indonesia, sarmini@amikompurwokerto.ac.id

ABSTRACT

Increasingly erratic rainfall has a significant effect on agriculture, and health. One field of agriculture that utilizes rainfall patterns is paddy fields. Rice farmers need a calculation of rainfall to determine the process of hatching and planting rice, especially for farmers who have rainfed rice fields. While in the health sector, rainfall greatly determines the growth of mosquitoes which can lead to dengue fever. The health office needs calculations to anticipate the number of people affected by dengue fever. In this study two data modeling were made in predicting rainfall in the Banyumas area, one data model or scenario using weather data other than rainfall such as temperature, humidity, wind speed, duration of solar radiation. While scenario two only uses historical rainfall data per day. In making models and testing data using the Artificial Neural Network (ANN) algorithm. The results showed that scenario two was better than scenario one after being evaluated using the Root Mean Square Error (RMSE). The RMSE obtained in scenarios one and two are 21,667 and 20,448, respectively. It can be concluded that to predict rainfall in the Banyumas area it is better to use historical rainfall data per day compared to using other weather data.

Key words: ANN, rainfall, predictions, historical data.

1. INTRODUCTION

The strategic geographical position of Indonesia, located in the tropics, between the Asian and Australian Continents, between the Pacific Ocean and the Indian Ocean, and through the equator, consists of islands and islands stretching from west to east, surrounded by vast oceans, causing Indonesian territory has a variety of weather and climate. Indonesia's climate diversity is influenced by global phenomena such as El Nino Southern Oscillation (ENSO) sourced from the Central Pacific Equator region and Indian Ocean Dipole (IOD) sourced from the western Indian Ocean region, Sumatra to east Africa, climate diversity is also influenced by regional phenomena, such as Asia-Australia monsoon wind circulation, the Inter Tropical Convergence Zone (ITCZ) which is a cloud growth area, and sea surface temperature conditions around the territory of Indonesia [1].

While the topographical conditions of the Indonesian region which have mountainous areas, buried, many beaches, are local topography which adds to the diversity of climate conditions in the Indonesian region, both according to space (region) and time. Based on the results of the average data analysis of the last 30 years (1981-2010), climatologically the Indonesian region has 407 climate patterns, of which 342 patterns are Seasonal Zones (ZOM), there are clear differences between the rainy season and the dry season, while 65 other patterns are Non-Zone (Non ZOM). Non-ZOM areas generally have 2 times the maximum rainfall in a year (the Equatorial pattern) or areas where throughout the year the rainfall is always high or low [1].

The high and low rainfall greatly affects the amount of water availability in the territory of Indonesia, where water is one of the vital needs for humans, in addition to daily activities or personal needs, water is also very needed by humans, especially Indonesian people who are mostly farmers, the yield of a farmer is largely determined by the amount of water that irrigates their agricultural land, the area of rice fields in Indonesia according to the Central Bureau of Statistics in 2015 is 8087393 ha, this area is very wide, so it requires a lot of water to irrigate farmers' fields. Not all paddy fields in Indonesia have good irrigation, there are many paddy fields that only depend on the availability of water for rain, for example in Banyumas Regency there are 6389 Ha of a total of 32115 Ha of rice fields whose water sources depend on rain [1]. So that the rice fields will dry out in the dry season and the production will drop. So it can be concluded that the amount of rice and rice production is very dependent on the availability of water [2][3].

In addition to agriculture affected by high and low levels of rain, the health sector is affected by the level of rainfall in an area. One disease that spreads based on rainfall level is the spread of dengue fever. The number of dengue cases has increased significantly around the world in recent decades [4][5]. Based on the latest data, the number of dengue cases per year is estimated to be around 390 million cases, areas that are prone to dengue fever cases are the tropics, namely America and Asia [5]. The Southeast Asian region has a periodic pattern of dengue fever, besides that a country has an

epidemic situation that has a lot of patterns[6]. The country that has the highest economic burden due to dengue fever is Indonesia [7].
In the context of deaths from dengue fever in Indonesia, as many as 1,229 people died in 2015 because of the disease caused by this dengue virus. In Banyumas, dengue is still a serious problem. The number of dengue cases in Banyumas Regency is caused by the existence of an unstable climate and quite high rainfall in the rainy season which is a potential breeding tool for Aedes Aegypty mosquitoes, also supported by the not yet optimal PSN activities in the community. Morbidity/Incidence Rate (IR) in Banyumas Regency in 2015 amounted to 13.6/100,000 population with 264 cases, when compared to 2014 amounted to 10.5/100,000 population then there was. The number of IRs in Banyumas Regency is still above the national target of <2/100,000 population, while the mortality rate/Case Fatality Rate (CFR) in 2015 is 1 person or 0.4% [8].
There are many studies that can predict the number of cases of dengue fever by utilizing weather data in an area, weather data used one of which is rainfall [9][10][11][12][13]. The study aims to provide an estimate of the number of cases in the future, so that the parties involved can anticipate cases of dengue fever in a particular area. However, there is a condition of season and climate uncertainty, which causes the error rate to increase in estimating rainfall which is the responsibility of meteorological service agencies throughout the world. The problem of estimating rain is usually approached in various ways. Satellite imagery can overcome the problem of rainfall prediction by capturing the current weather conditions, this can be done for short-term estimates. However, this method still creates a problem, namely the inaccuracy of rainfall predictions.
One other way to get the rainfall forecasting model is by time series analysis techniques. There are several different approaches to time series forecasting, especially for climate proposals. Traditional forecasting has long been a domain of linear statistics, an approach commonly used for prediction of time series, such as Box-Jenkins [14] or ARIMA (autoregressive integrated moving average) method, assuming that the time series behaves as a process linear. Regardless of understanding and easy application, it may really be inappropriate to apply if the ongoing mechanism undergoes a nonlinear process.
Recently researchers have approached such problems with artificial neural networks (ANN), this is a strong alternative to traditional time-series modeling [15] as for the NWP model. ANN is a data-based self-adaptive method that is able to understand and solve problems that do not have enough data or observations to use more traditional statistical models [15], rainfall is such a phenomenon and ANN is suitable and studied solutions.
In this study, we will apply ANN with two different experiments to predict rainfall. The first experiment uses rainfall data and other8suppornhng data as input attributes, while the second experiment uses only rainfall data. The results of this prediction will be used to anticipate water needs in agriculture and the spread of dengue fever in the Banyumas district.

2. RELATED WORKS

The linear, conceptual, and statistical models are the basic models used to forecast rainfall since a number of periods ago. ARIMA is the most popular model in predicting rainfall [16][17]. By utilizing 31 rainfall stations in Thailand, Research [18] predicts rainfall with ARIMA. While Research [19] predicts rainfall in Dhaka Bangladesh on a monthly basis by utilizing ARIMA.
Artificial Neural Network is a method that is often used in the last two periods in predicting rainfall [20][21][22], in addition ANN is also often used for forecasting reservoir inflow [23], total annual crude oil export [24], and prediction of sea level [23]. Based on the research that has been done, it can be concluded that ANN is the right method to succeed and predict. ANN can overcome challenges with ever increasing increases [25][26].

3. METHOD

3.1 Data

The data used in this study is weather data taken from BMKG Cilacap, the data taken is weather data from January 2010 to June 2019. Table 1 shows the data structure used in this study. While Table 2 shows the data description in Table 1.

Table 1: Examples of data used in this study

Tn	Tx	Tavg	RH_avg	ss	ff_x	ddd_x	ff_avg	RR
24,8	32,5	28,8	83	6,1	8	270	1	5,4
26,2	32,8	29,5		6,6	7	210	2	11,3
	32,6	29,5	74	3,9	6	120	2	0
26,2	32,7		78	6,9	7	130	2	8888
24,4	32,6	27,6	82	4,7	6	320	2	56,3

Table 2: Description of attributes in the data used

Atribut	Information
8888	: Data is not measurable
Tn	: Minimum temperature (°C)
Tx	: Maximum temperature (°C)
Tavg	: Average temperature (°C)
RH_avg	: Average humidity (%)
RR	: Rainfall (mm)
ss	: The duration of solar radiation (jam)
ff_x	: Maximum wind speed (m/s)
ddd_x	: Wind direction at maximum speed (°)
ff_avg	: Average wind speed (m/s)

Based on Table 1. There are empty and unmeasured data. This happened due to an error during data retrieval carried out by BMKG Cilacap, even though the retrieval process was carried out using radar or satellite, but there were still opportunities for errors. To handle empty data, we use the average for each attribute to fill in the blank data [27]. Likewise, on unmeasured data, the data is considered empty and is replaced by the average of the attribute.

In this study in predicting rainfall using two scenarios, the first scenario is to use data in Table 1, where Rainfall (RR) is used as Class/Label while the other is used as an attribute/feature. The second scenario only uses rainfall data, three days are used as attributes/features and the next day is used as class/label. Table 3 shows the data in the second scenario.

Table 3:3 Data pada skenario kedua

Atribut 1	Atribut 2	Atribut 3	Class
5.4	11.3	0	8888
11.3	0	8888	56.3
0	8888	56.3	0
8888	56.3	0	8888
56.3	0	8888	44.2
0	8888	44.2	0
8888	44.2	0	79.9
44.2	0	79.9	12

Setting parameters can play an important role to get the desired results. Although there are no fixed rules that must be followed in setting these parameters, in general, the choice still depends on the type and size of the training dataset [28]. In making a model to predict rainfall, this study uses data from 2010 - 2015, while to test the resulting model, the data used is from January 2016 - June 2019. The distribution of the data is used in the first scenario and the second scenario.

3.2 Artificial Neural Network

Based on the research that has been done, Artificial Neural Network is an efficient algorithm, the algorithm is able to understand system behavior well and effectively [29]. Multi-layer Artificial Neural Network is an adaptive method that can change its structure to solve problems based on external and internal information flowing through the network. Artificial Neural Networks can be used to model complex relationships between input and output values to find patterns in the data. Because it was inspired by the working model of the brain's biological neural networks, Artificial Neural Networks process large amounts of information in parallel and distributed [30][31][32][33][34].

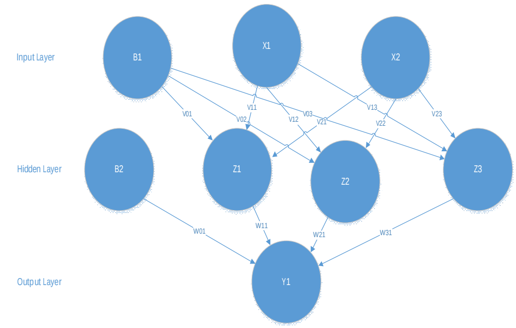

Figure 5: Arsitecture of Artificial Neural Network

Multi-layer Artificial Neural Network Architecture has at least 3 layers, namely input layer, hidden layer, and output layer. Each layer has at least one neuron. A simple example of a multi-layer Artificial Neural Network architecture is shown in Figure 5. From Figure 5 it is known that variables B1 and B2 are the biases of neurons, X1 and X2 are input neurons, Y1 is the output of neurons, Z1, Z2, and Z3 are neuronal hidden while the line between neurons is the weight of neurons.

3.3 Performance Evaluation Criteria

The error calculation used in this study uses a calculation called the Root Mean Square Error as in Equation (1) where y is the error obtained, x 'is the result that has been obtained, x is the result that should be obtained, i is the sequence of data practice, and n is the amount of training data.

$$y=\sqrt{\frac{\sum_{i=1}^{n}(x'_i - x_i)^2}{n}} \qquad (1)$$

4. RESULT AND DISCUSSION

In this study predictions of rainfall in the Banyumas area with historical data obtained from BMKG Cilacap. The amount of data used to make the model in scenario 1 and scenario 2 is 1966, while the data used to test the model obtained is 1122. Tables 4 and 5 show the results of statistical data used in scenarios 1 and 2 respectively.

Table 4: Data statistics in scenario 1

	Min	Max	Average
RR	0	199.5	15.553
Predictions (RR)	-122.226	35.349	15.041
Tn	19.8	27.1	24.771
Tx	3.6	34.7	31.156
Tavg	24.2	30.5	27.448
RH_avg	66	97	83.492
ss	0	99	6.555
Ffx	1	12	4.79
Dddx	10	360	168.148
Ff_avg	0	6	1.868

Table 5: Data statistics in scenario 2

	Min	Max	Average
RR	0	199.5	15.553
Predictions (RR)	18.313	59.320	21.492
RR1	0	199.5	15.905
RR2	3.6	1995.5	15.891
RR3	24.2	1995.5	15.903

Based on Table 4 and Table 5 there are additional variables named Predictions (RR), the variable is the variable; which shows the results of rainfall predictions during the process of testing the model with test data. The results of these predictions can be seen in detail in Table 6 and Table 7, the predicted results will be compared with actual rainfall data.

Table 6: Rainfall prediction results in scenario 1

RR	Predictions (RR)	Galat
0	7.366	7.366
4.4	5.709	1.309
7.309	6.449	0.86
....	....	....
3.5	9.203	5.703
7.309	10.438	1.298
7.309	6.011	1.298
0	3.069	3.069

Table 7: Rainfall prediction results in the scenario 2

RR	Predictions (RR)	Galat
7.309	18.608404665305088	11,299
7.309	18.978294214101354	11,669
7.309	19.312798253714945	12,003
....	....	....
7.309	18.945518360537235	11,636
38.8	19.164874720087923	19,636
3.5	21.646587487024107	18,146
7.309	20.821362655245565	13,512

Prediction results that have been obtained from two scenarios that have been designed, then evaluated using the Root Mean Squared Error (RMSE). Table 8 shows the RMSE comparison of scenario 1 and scenario 2.

Table 8: Comparison of RMSE results from scenario 1 and 2

Skenario	RMSE
1	21.667
2	20.478

Based on Table 8, scenario two has a lower error rate compared to scenario one, in other words the prediction of rainfall in the Banyumas area is better to use previous rainfall compared to other weather data.

5. CONCLUSION

ANN algorithm can be used to predict rainfall in the Banyumas area using either scenario one where rainfall is predicted using other weather data and scenario two where rainfall is predicted with previous rainfall data. Based on the results of tests conducted in which the evaluation was measured by RMSE, it was found that the accuracy of rainfall predictions using scenario two was better than scenario one. Although the difference in accuracy between scenarios one and two is not so large, it has been proven that predicting rainfall in the Banyumas area in the future is better based on previous rainfall data.

ACKNOWLEDGEMENT

We thank the LPPM Universitas Amikom Purwokerto for providing funding for this study through the Research Grant Program for Universitas Amikom Purwokerto.

REFERENCES

1. Badan Pusat Statistik Indonesia, **Stastical Yearbook of Indonesia 2018**, Jakarta, 2018.
2. L. J. Wade, S. Fukai, B. K. Samson, A. Ali, and M. A. Mazid, **Rainfed lowland rice: Physical environment and cultivar requirements**, *F. Crop. Res.*, vol. 64, no. 1–2, pp. 3–12, 1999. https://doi.org/10.1016/S0378-4290(99)00047-7
3. D. C. Little, P. Surintaraseree, and N. Innes-Taylor, **Fish culture in rainfed rice fields of northeast Thailand**, *Aquaculture*, vol. 140, no. 4, pp. 295–321, 1996.
4. S. Bhatt *et al.*, **The global distribution and burden of dengue**, vol. 496, no. 7446, pp. 504–507, 2013. https://doi.org/10.1038/nature12060
5. S. Bhatt *et al.*, T**he global distribution and burden of dengue**, *Nature*, vol. 496, no. 7446, pp. 504–507, Apr. 2013. https://doi.org/10.1038/nature12060
6. A. S and L. P, **Emerging trends of malaria-dengue geographical coupling in the Southeast Asia region**, *J. Vector Borne Dis.*, vol. 51, no. 3, pp. 165–171, 2014.
7. D. S. Shepard, E. A. Undurraga, and Y. A. Halasa, **Economic and Disease Burden of Dengue in Southeast Asia**, *PLoS Negl. Trop. Dis.*, vol. 7, no. 2, p. e2055, Feb.

2013.
https://doi.org/10.1371/journal.pntd.0002055

8. Dinas Kesehatan Kabupaten Banyumas, **Profil Kesehatan Kabupaten Banyumas Tahun 2015**, Banyumas, 2015.
9. A. M. Najar, M. I. Irawan, and D. Adzkiya, **Extreme Learning Machine Method for Dengue Hemorrhagic Fever Outbreak Risk Level Prediction**, *2018 Int. Conf. Smart Comput. Electron. Enterp. ICSCEE 2018*, pp. 1–5, 2018.
10. G. Sánchez-González, R. Condé, R. N. Moreno, and P. C. López Vázquez, **Prediction of dengue outbreaks in Mexico based on entomological, meteorological and demographic data**, *PLoS One*, vol. 13, no. 8, pp. 1–14, 2018.
11. C. Hettiarachchige, S. von Cavallar, T. Lynar, R. I. Hickson, and M. Gambhir, **Risk prediction system for dengue transmission based on high resolution weather data**, *PLoS One*, vol. 13, no. 12, pp. 1–17, 2018.
12. H.-Y. Yuan *et al.*, **Prediction of Annual Dengue Incidence by Hydro-Climatic Extremes for Southern Taiwan**, *Int. J. Biometeorol.*, 2019.
13. A. L. Ramadona, L. Lazuardi, Y. L. Hii, Å. Holmner, H. Kusnanto, and J. Rocklöv, **Prediction of dengue outbreaks based on disease surveillance and meteorological data**, *PLoS One*, vol. 11, no. 3, pp. 1–18, 2016.
https://doi.org/10.1371/journal.pone.0152688
14. G. E. P. Box, G. Jenkins, G. C. Reinsel, and G. M. Ljung, **Time series analysis: forecasting and control**. John Wiley & Sons, 2015.
15. P. G. Zhang, **Time series forecasting using a hybrid ARIMA and neural network model**, *Neurocomputing*, vol. 50, pp. 159–175, 2003.
16. B. S. Kim, S. Z. Hossein, and G. Choi, **Evaluation of temporal-spatial precipitation variability and prediction using seasonal ARIMA model in Mongolia**, *KSCE J. Civ. Eng.*, vol. 15, no. 5, pp. 917–925, 2011.
17. I. Kaushik and S. Singh, **Seasonal Arima Model for Forecasting of Monthly Rainfall and Temperature**, *J. Environ. Res. Dev.*, vol. 3, no. 2, pp. 506–514, 2008.
18. U. Weesakul and S. Lowanichchai, **Rainfall forecast for agricultural water allocation planning in Thailand**, *Thammasat Int. J. Sc. Tech*, vol. 10, no. 3, pp. 18–27, 2005.
19. M. Mahsin, Y. Akhter, and M. Begum, **Modeling Rainfall in Dhaka Division of Bangladesh Using Time Series**, *J. Math. Model. Appl.*, vol. 1, no. 5, pp. 67–73, 2012.
20. J. T. Esteves, G. de Souza Rolim, and A. S. Ferraudo, **Rainfall prediction methodology with binary multilayer perceptron neural networks**, *Clim. Dyn.*, vol. 52, no. 3–4, pp. 2319–2331, 2019.
21. T. Kashiwao, K. Nakayama, S. Ando, K. Ikeda, M. Lee, and A. Bahadori, **A neural network-based local rainfall prediction system using meteorological data on the Internet: A case study using data from the Japan Meteorological Agency**, *Appl. Soft Comput. J.*, vol. 56, pp. 317–330, 2017.
22. E. G. Wahyuni, L. M. F. Fauzan, F. Abriyani, N. F. Muchlis, and M. Ulfa, **Rainfall prediction with backpropagation method**, *J. Phys. Conf. Ser.*, vol. 983, no. 1, 2018.
https://doi.org/10.1088/1742-6596/983/1/012059
23. I. Rasooli, M. L. Moosavi, M. B. Rezaee, and K. Jaimand, **Comparison Of Regression, Arima And Ann Models For Reservoir Inflow Forecasting Using Snowmelt Equivalent (A Case Study Of Karaj)**, *J. Agric. Sci. Technol.*, vol. 4, pp. 17–30, 2002.
24. T. Alam, **Crude Oil Export: A Comparative Analysis using Artificial Neural Network and ARIMA**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 5, pp. 2546–2550, 2019.
https://doi.org/10.30534/ijatcse/2019/102852019
25. A. Kohli and P. Arora, **Application of Artificial Neural Networks for Well Logs**, pp. 27–32, 2016.
26. A. M. Khidir, H. H. A. Adlan, and I. A. Basheir, **Neural Networks forecasting architectures for rainfall in the rain-fed Sectors in Sudan**, *Proc. - 2013 Int. Conf. Comput. Electr. Electron. Eng. 'Research Makes a Differ. ICCEEE 2013*, pp. 700–707, 2013.
https://doi.org/10.1109/ICCEEE.2013.6634026
27. W. M. Baihaqi, N. A. Setiawan, and I. Ardiyanto, **Rule extraction for fuzzy expert system to diagnose Coronary artery disease**, in *Proceedings - 2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2016*, 2016.
28. M. Akour, O. Al Qasem, H. Alsghaier, and K. Al-radaideh, **The Effectiveness of Using Deep Learning Algorithms in Predicting Daily Activities**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 5, pp. 2231–2235, 2019.
https://doi.org/10.30534/ijatcse/2019/57852019
29. J. Zurada, **Introduction to Artificial Neural Systems**. St. Paul, MN, USA: West Publishing Co., 1992.
30. S. Haykin, **Neural Network A Comprehensive Foundation,** 2 edition. Prentice Hall, 1998.
31. T. Hariguna, & Berlilana, "Understanding of Antecedents to Achieve Customer Trust and Customer Intention to Purchase E-Commerce in Social Media, an Empirical Assessment", International Journal of Electrical and Computer Engineering, 7(3), 1240–1245, 2017. http://doi.org/10.11591/ijece.v7i3.pp1240-1245
32. T. Hariguna, CW. Hung, & H.T. Sukmana, "The antecedent of citizen intention use of e-government service, TELKOMNIKA (Telecommunication Computing Electronics and Control)", 17(1), 202–209, 2019.
http://doi.org/10.12928/TELKOMNIKA.v17i1.11588
33. U. Rahardja, T. Hariguna, & Q. Aini, "Understanding the Impact of Determinants in Game Learning Acceptance: An Empirical Study", International Journal

of Education and Practice, 7(3), 136–145, 2019. http://doi.org/10.18488/journal.61.2019.73.136.145

34. R. Nanda, “Effective Analysis of Human Facial Appearance using JAFFE Images,” Int. J. Adv. Trends Comput. Sci. Eng., vol. 7, no. 6, pp. 149–151, 2018. https://doi.org/10.30534/ijatcse/2018/17762018