The Effect of Number of Factors and Data on Monthly Weather Classification Performance Using Artificial Neural Networks

Shofura[1] , Sri Suryani Prasetiyowati *[2], Linda Salma A[3], Sri Harini [4]

[1,2] *School of Computing, Telkom University*
Bandung, Indonesia

[3]*Jacobs University*
Bremen, Jerman

[4]*Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim*
Malang, Indonesia

shofuraf@student.telkomuniversity.ac.id, *srisuryani@telkomuniversity.ac.id,
L.angreani@jacobs-university.de, sriharini@mat.uin-malang.ac.id

Abstract

There are two seasons in Indonesia, namely the rainy season and the dry season, which alternates every 6 months. In information technology, information regarding the prediction of weather classification in the next few months is very needed, and can be used for decision makers, namely the government and the public. This study examines the effect of the number of factors and the amount of data on the performance of the prediction of monthly weather classifications. Performance comparisons are made based on weather predictions based on 4 factors as used in previous research, namely: average temperature, solar radiation, air pressure, wind, against predictions based on 5 factors, namely with an additional factor of rainfall. While the comparison of the amount of data, the amount is based on the range of years used in the classification process. The classification prediction model used in this study is the Artificial Neural Network (ANN) method combined with the backpropagation algorithm to calculate the weight of the ANN network. The data used is meteorological data for 8 years, in the period 2010-2018. The experiment results showed that the design of the classification prediction model using the amount of data in a 6-year period with 5 factors has an accuracy rate of 83.33%. This value is higher than using 4 factors. However, in the ANN classification model using 4 factors, the accuracy is more stable even though the amount of data varies. By using analysis of variance it can be seen that the number of factors will have a significant effect on the weather, with a confidence interval of 84%, while the use of the amount of training data will have a significant effect on the accuracy of the weather classification results, with a confidence interval of 92%.

Keywords: backpropagation, artificial neural networks, predictions, monthly weather, rainfall

I. INTRODUCTION

CLIMATE is the average weather conditions in the long term and covers a wider area. Indonesia is located in a tropical climate which is crossed by the equator which is located at 23.5°N-23.5°S. Tropical climates only have 2 seasons, namely the dry season and the rainy season[1]. The impact of global warming is the

occurrence of climate change which in turn will affect life on earth [2]. Climate change is one of the issues that is currently being discussed, because its impact greatly affects all aspects of human life [1]. Climate can be determined based on weather conditions in each month. Monthly weather prediction is a challenging issue in the climatology department in recent years. Even after advances in technology, accuracy in predicting climate has never been high [3].

The territory of Indonesia is an area that is influenced by monsoon climatic conditions, namely there is a clear difference between wet season and dry season. Climate change is marked by changes in rainfall patterns, which causes a shift in the beginning of the planting season so that it is difficult to make plant cultivation plans [4]. This can lead to crop failure of agricultural products, resulting in significant losses. Therefore, accurate information is needed regarding the climatic conditions (rainfall) of an area in the past and at present, to minimize these losses [2].

Accurate predictions of weather or rainfall are needed, because weather and rainfall are nonlinear and dynamic problems. Therefore, machine models and simulations are needed for accurate predictions. In Indonesia, agriculture is one of the most important economic fields, so increasing agricultural production is one of the government's priority programs. For farmers, accurate prediction of rainfall is very important to know the right planting season so that the productivity of agricultural activities increases. Because the factor of sufficient water availability will support agricultural productivity. So it is necessary to involve science and technology to determine weather conditions [5].

During an energy crisis, wind power is a significant alternative energy. Due to the intermittent and fluctuating it can be an operational challenge in grid-connected wind energy systems. Today wind speed forecasting has become one of the most interesting topics in the field of renewable energy, because it can produce clean energy, and its capacity can be integrated into the grid. [6] and [7] research compared the autoregressive integrated moving average (ARIMA) model, with artificial neural networks (ANN), RNN and LSTM, to estimate wind speed in the future. The model is applied to wind speed data for each month. The study aims to find the most effective predictive model on time series, with better accuracy. [6] experimental results show that the ANN model does a better job than the ARIMA model. Meanwhile, [7] research shows that the LSTM method is more accurate than ARIMA, but without showing the results of ANN and RNN.

Similarly, according to [8] research on a review of the methods used in rainfall research, which was carried out by comparing the results of studies using the Fuzzy model, Fast Fourier Transformation (FFT), Emotional Artificial Neural Network (EANN), Artificial Neural Network (ANN) , Adaptive Ensemble Empirical Mode Decomposition-Artificial Neural Network (AEEMD-ANN), E-SVR-Artificial Neural Network (E-SVR-ANN), Artificial Neural Network Backpropagation (BPNN), Adaptive Splines Threshold (ASTAR), Seasonal First-Order Autoregressive (SAR), Gumbel, Autoregressive Integrated Moving Average (ARIMA), Feed Forward Neural Network (FFNN), Support Vector Machine (SVM), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Artificial Neural Network-Fuzzy (ANN Fuzzy). The results of the review state that the Artificial Neural Network model is superior to other methods in recognizing patterns well and is easier to develop according to existing problems and parameters. So that ANN is recommended as a method used in research on rainfall prediction.

Based on these studies, in this study, the effect of using the number of factors and data on the weather prediction process using the Artificial Neural Network classification method will be studied. This method was chosen because based on [6], [7] research, ANN is a recommended method for studies related to wind speed and rainfall and produces a high level of accuracy [7], [8]. In this study, there is an additional weather factor, namely rainfall in the previous month. The study of the effect of the number of variations of factors and data on the performance of weather prediction classification results using ANN, shows the contribution of this study compared to previous studies.

II. LITERATURE REVIEW

Data mining is one of the most frequently used techniques to predict climate or weather where one of the tasks in data mining is a classification method. In this classification method there are various classifier methods that can be used such as Naïve Bayes Classifier, Support Vector Machine [9], Decision Tree Algorithm C4.5 [10], Artificial Neural Networks (ANN) [11] etc. These studies concluded that many studies have used ANN in predicting future weather and also concludes that data mining techniques, especially ANN can be used to predict weather conditions accurately.

[5] states that some researchers use ANN for rainfall prediction, because it is a valid and more accurate method than conventional mathematics or numerical approaches. This paper discusses the comparison of the predicted perceived value with BPN, RBFN, SVM. The factors used to predict rainfall are humidity, air pressure, and temperature. The results obtained are the highest perception value of 0.93% issued by SVM predictions, this value is greater than that produced by other neural network methods, namely BPN and RBFN. This paper does not explain the analysis in detail, the basis of the perception calculation used. Meanwhile, according to [12], numerical weather prediction model (NWP) or statistical model which is a traditional method, cannot provide a significant influence on rainfall prediction, because its characteristics are nonlinear and dynamic. While ANN can solve nonlinear relationships between variables, so it is suitable for predicting rainfall. In this study, a combination of ANN and several algorithms using a neural network for rainfall prediction is combined, so that accuracy can increase rapidly. Similarly, the results of [13] reviews of rainfall prediction studies conducted by various authors using the Artificial Neural Network technique. Back-Propagation, Auto-Regressive Moving Average (ARIMA), ANN , K-Nearest Neighborhood (K-NN), Hybrid Model (Wavelet ANN), Wavelet-NARX Hybrid Model, Rainfall Model, Two-stage optimization technique, Adaptive Basis Function Neural Network (ABFNN), Multilayer perceptron, etc., most of them state that the accuracy is more than 95%, the results of rainfall prediction using ANN techniques are far superior to other techniques such as Numerical Weather Prediction (NWP) and Statistical Methods, due to physical conditions non-linear and complex that affect the occurrence of rainfall.

Research on the application of data mining for weather prediction with a focus on climate change prediction. was also carried out by [10], [14]. The ANN method and decision tree C5 are used as the classifier model. The weather factors used in these two studies are maximum temperature, minimum temperature, rainfall, evaporation and wind speed. The experimental results show that both classifier models can be used as weather prediction models, but the error generated by the decision tree model is still quite large, which is more than 40%. The resulting C5 model is also still not simple so it requires further action to obtain a simpler model. The ANN model gives a lower prediction error than C5, but the resulting error is not too small, which is more than 20%.

Implementation of the classification method to classify thirteen different plant species is conducted by [15] where three methods i.e. SVM, ANN and CNN were used. Accuracy is obtained successively, CNN is 99%, ANN is 94%, and SVM is 91%. Each classifier is also tested with scenarios of increasing the number of samples for training and segmentation size. Increasing training samples improves SVM performance. Based on the accuracy results obtained, it can be said that all classification methods provide very high classification accuracy, with CNN being 5% superior to ANN.

Meanwhile, in [16] which developed a rainfall prediction model in late spring-early summer, using a neural network (ANN), selected 11 significant input variables for the initial structure of ANN. The attributes used are the East Atlantic (EA) lagging climate index pattern, North Atlantic Oscillation (NAO), Pacific Decadal Oscillation (PDO), East Pacific Oscillation/North Pacific (EP/NP), and Tropical Northern Atlantic (TNA) Index. With these five input variables, the best ANN model shows performance as measured by the relative root mean square error for each training, validation and test data of 25.84%, 32.72%, and 34.75%. Based on the hit score, where the number of hit years divided by the number of years, it was found that more than 60%, the ANN

model was successful in predicting rainfall in the area, thus enabling more timely and flexible water resource management. Thus the potential for drought in the region can be resolved.

Research by [17] on climate models using artificial neural networks (ANN) was applied to monthly temperature and rainfall data for the base time (1979–2009) at four different metrology stations, the future temperature and annual mean rainfall were predicted to be up to 2100 Large-scale GCM predictors were created under scenarios A1B and A2 until the 21st century. This study it is said that artificial neural networks (ANNs) are analogous to multiple regression, to overcome nonlinear data and noisy data. In this research, double ANN feed forward based on layered perceptron using back propagation learning algorithm is used. Two types of transfer functions are; pure-line and tan-sig as transfer functions, with the first two networks consisting of an input layer with four input parameters, and a hidden layer consisting of nine neurons, one output. While the last two networks consist of an input layer with ten input parameters, two hidden layers with nine neurons in the first hidden layer and one neuron in the second layer. Meanwhile [18], to predict the level of rainfall is very complex with a large number of parameters, using an alternative approach based on time-series models. One of the algorithms that is widely used to predict the future is Neural Network Backpropagation. Nguyen-Widrow method is used to initialize the Neural Network weights and reduce training time. In addition, use maximum epoch 50 and 3 neurons in hidden layer.

Research related to weather prediction has also been carried out by a number of researchers [3], [9], [11]. Weather referred to here includes temperature, humidity, wind, rain, and climate. Based on [11] it is concluded that there are 4 main factors used in predicting the weather, namely temperature, rainfall, average length of irradiation and wind speed. Especially for the research on rain prediction conducted by [9], the average factor used is 4 factors. The rainfall factor in the previous month has not been studied in the 3 studies. In all the research related to weather prediction, the modeling of training data was never discussed, even though if the training data used in the study came from the same time period, the use of the classification method failed to be used as a predictive model.

Multi-layered artificial neural network with back-propagation algorithm configuration according to this research is the most commonly used, because of the ease of training. In this study, training and testing data were built, in order to find the number of neurons are hidden in this layer, resulting in the best performance. In study by [19], predictions of the average rainfall in Udupi district of Karnataka with an artificial neural network model, where the predictive model is based on three layered networks and the number of hidden neurons is different. The three algorithms tested on the multi-layer architecture are Back Propagation Algorithm (BPA), Layer Recurrent Network (LRN), and Cascaded Back-Propagation (CBP). From the experimental results, it is found that when the number of neurons increases in the ANN, the Mean Square Error (MSE) decreases. Meanwhile, [20] research uses back propagation in its neural network because it is easy to train and has accuracy. The algorithm consists of two aspects, namely to generate network input patterns and adjust the output through changes in network weights. Back propagation algorithm can be used to predict rainfall. Rainfall predictions were analyzed using a back propagation neural network algorithm, which is a three-layer model to train and study different attributes of hidden neurons in the network.

Based on the studies, most papers stated that ANN is superior to other methods, so this study will use of the ANN method to obtain a weather classification model based on the factors that influence it. In general, ANN is an information processing paradigm that is inspired by the way the biological nervous system works, such as how the brain processes information [11]. ANN can be classified into 2 categories based on the type of connection between neurons, namely feed forward and backpropagation, but in this study, the algorithm used is backpropagation. Backpropagation is a type of controlled training that uses a weight adjustment pattern to achieve a minimum error value between the predicted output and the real output. In designing the system, this study uses two scenarios, namely comparing the effect and assessing the trend of the accuracy results based on the number of factors and the amount of different data.

III. RESEARCH METHOD

The stage of weather prediction research using this artificial conditional network start with the data modeling process, where after this stage data modeling is carried, that the data is ready to be used for the next process. The next stage is preprocessing which includes the cleaning process and data transformation. The last process is ANN model training which aims to obtain the most optimal ANN model. Several variations of the data model are used to obtain the most optimal model.

A. Dataset

The data used in this study is daily weather data in Bogor which comes from the Darmaga station of the Meteorology, Climatology and Geophysics Agency (BMKG). The dataset used includes daily weather data from 2010 to 2018 with a total of 3287 data for the daily scale and 108 data for the monthly scale data. Daily weather data comes from the head office of the Meteorology, Climatology and Geophysics Agency (BMKG) with factors or factors that are relevant to climate predictions such as average temperature, duration of sunshine, rainfall, air pressure and wind speed. TABLE I shows an explanation of the units and scales for each factor used. It can be seen from each factor has a different unit of measurement.

TABLE I
WEATHER FACTORS

Factors	Unit	Scale
Average temperature	Temperature Scale	Celsius
Sun exposure time	Percentage of duration of irradiation	Percentage
Rainfall	Unit length	Millimeter
Air pressure	Metric unit of pressure	M-bar
Wind velocity	Unit of speed	Km/Hour

The data obtained from BMKG is data on a daily time scale, while in this study data is needed on a monthly scale. Therefore, the daily weather data is converted into a monthly scale. Data on the rainfall factor can be used as a reference in determining the occurrence of rain. Rainfall at number 0 can be stated as no rain, for rainfall with number 8888 it can be stated that rain is not measured or there is rain but very little so it can be considered 0. So, the daily rainfall data will be converted into daily weather data basic and monthly, for the number 8888 will be replaced with the number 0.

TABLE II
MONTHLY WEATHER DATA SAMPLE

i Month	Average Temperature	Sun exposure time	Rainfall	Air Pressure	Wind Velocity
1	25.3	38.3	255	1012.6	2
2	25.9	59.2	463	1012.4	1.8
3	26	56.2	419	1012.1	2.7
4	27	78.4	43	1011.5	2.6
5	26.7	71.8	333	1010.2	2.7

Conversion of data into monthly scale data is carried out by means of daily weather data for the temperature scale, duration of sun exposure, air pressure and wind speed taken on a per-basis basis for each factor. As for the rainfall factor, it is accumulated per 30 days with a note that the number 8888 in the daily rainfall data is

replaced with the number 0 before it is accumulated and converted into basic scale data. TABLE II shows a sample for monthly weather data obtained from the daily data conversion results.

The data set used is divided into 2 parts, namely the data set for model training and the data set for prediction. Before predicting the climate that will occur, you must process the training data to get the best model based on a small error value. To get a good and accurate model, it is necessary to optimize the training data. The training data used is data from 2010-2016 and the test data uses data from 2017-2018.

After the data set is converted into 3 data in different time scales, the next process is labeling. Labeling is divided into 2 labels, namely "rain" and "no rain". Labeling is done by calculating the accumulation of rainfall from monthly weather data on rainfall factor data. If in one month the amount of rainfall is above 150 mm or more then it can be stated as "rain", but if in one month the amount of rainfall is below 150 mm then it can be declared "no rain". For example, in the Bogor area, for monthly data in January 2010 the number is 255 mm (above 150 mm), it can be stated that January has entered the rainy season. The labeling of weather data is carried out in the following month based on the level of rainfall for a month. An example of the results of labeling this data can be seen in TABLE III. In the example of TABLE III, the rainfall information for the X-month is used to label the X+1 month's weather. For example, for the second row of rainfall of 130, the weather label in the 1st row is "No rain".

TABLE III
RESULT OF LABELING WEATHER X+1 MONTH

Weather i+1 Month	Average Temperature	Sun exposure time	Rainfall	Air Pressure	Wind Velocity
Not Raining	25.3	38.3	255	1012.6	2
Rain	25.9	59.2	130	1012.4	1.8
Not Raining	26	56.2	419	1012.1	2.7
Rain	27	78.4	43	1011.5	2.6
Rain	26.7	71.8	333	1010.2	2.7

The training data modeling used in TABLE III shows that the weather prediction model to be built is the X+1-month weather prediction model based on weather factors in month X. This model can be developed based on the desired prediction model, for example based on data 2 previous month and so on. The use of all data components in the same time period will not be able to produce a predictive model, but can only be used to determine important weather factors in one-time period.
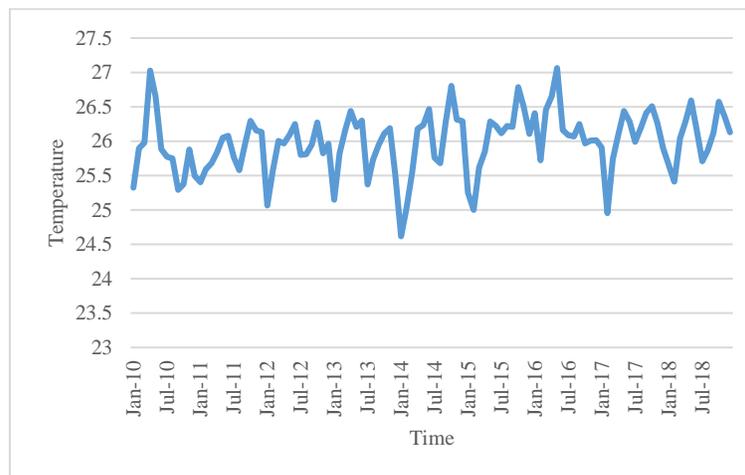

Fig. 1. Plot of Average Temperature from 2010-2018

The plot of each factor that is used as input from the ANN is shown in the Fig. 1 to Fig. 5. In Fig. 1., we can see that there is an up and down trend, this is because the average temperature changes according to the seasons. When the season is dry season, the minimum temperature will tend to be higher than the rainy season.
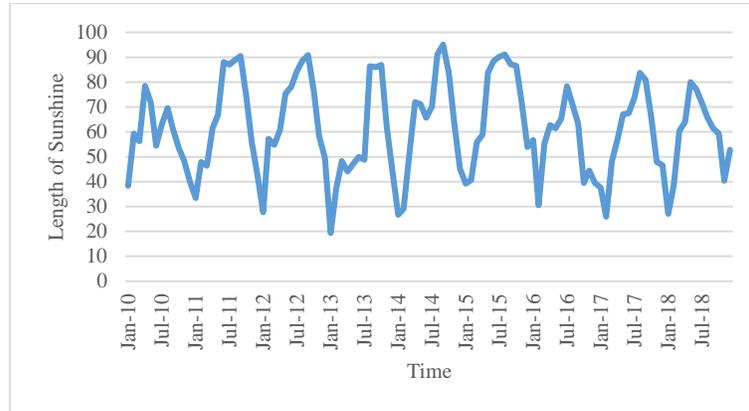

Fig. 2. Plot of Average Sunshine Time Duration from 2010-2018

While the pattern of the average sunshine time duration in Fig. 2. tends to have the same pattern within a certain period of time. In Fig. 3 Rainfall patterns appear to have different patterns in each certain period of time. The highest level of rainfall occurred in the period 2011-2012. Meanwhile, very low rainfall occurred in the period 2014-2015.
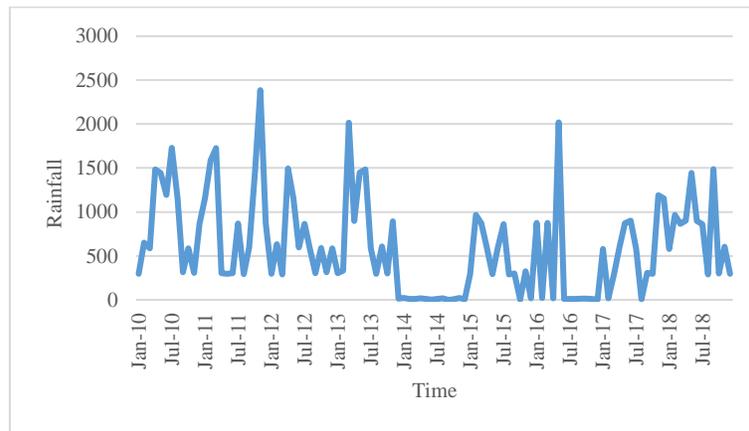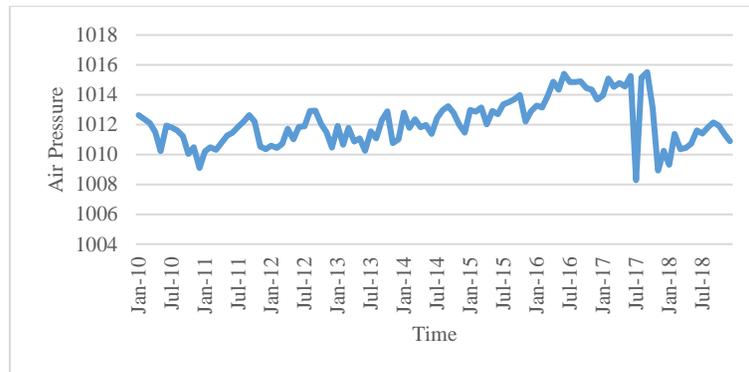

Fig. 3. Plot of Average Rainfall from 2010-2018

Fig. 4. Plot of Average Air Pressure from 2010-2018

Fig.4 show air pressure pattern that occurred from 2010-2018. Seen from 2010-2017, air pressure has an increasing trend. However, a different pattern occurred in the 2017-2018 period, where there was a fairly large and fluctuating decrease in air pressure, which the pattern then rose again in early 2018. A different pattern is also shown in Fig 5. where the wind speed level tends to decrease in the 2010-2011 time period, then the trend changes to increase in the 2011-2018 period. The anomaly occurred at the end of 2013, where the wind speed decreased drastically.
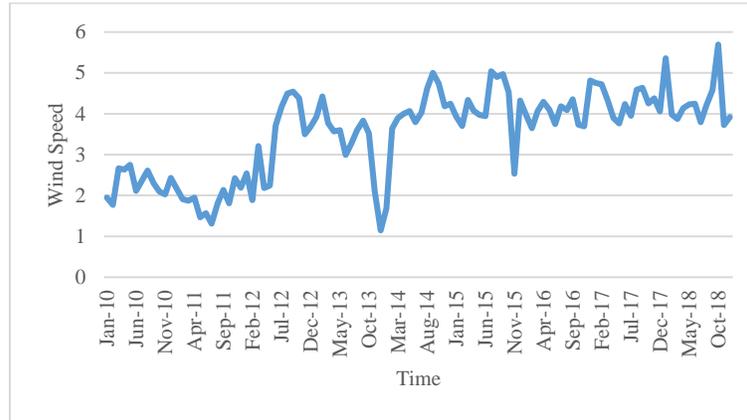

Fig. 5. Plot of Average Wind Speed from 2010-2018

B. Preprocessing

The data obtained from the BMKG is still in the form of original data from observations and there are still some data that are empty. Preprocessing is done to make the original data into quality data, so that it can be used in research. At this stage, data cleaning will be carried out first. The next process is data transformation, which is normalizing the data so that the data range becomes small. The normalization method used in this study is to transform the data into a range of 0 to 1 [21]. In Neural Networks, normalization of input data can improve network performance and reduce errors in the training process [22]. The normalized data equation used can be seen in equation (1).

$$X' = \frac{0.8\,(X - b)}{(a - b)} + 0.1 \tag{1}$$

Where,

   X' = Normalization Data

   X = Original Data

   a = Maximum value of original data

   b = Minimum value of original data

C. ANN Model Training

The classification model for prediction used in this study is the artificial neural network method, which is an information processing paradigm that is inspired by the workings of the biological nervous system of the human brain in processing information [11]. ANN is believed to be able to solve complex problems having a non-linear relationship [22]. The five input factors used are the average temperature (X_1), Length of sunshine (X_2), rainfall (X_3), air pressure (X_4), and wind speed (X_5). The number of hidden-layers in this study uses 11 hidden-layers using the formula (2N+1) [22]. The ANN architecture in this study can be visualized as shown in Fig. 6.
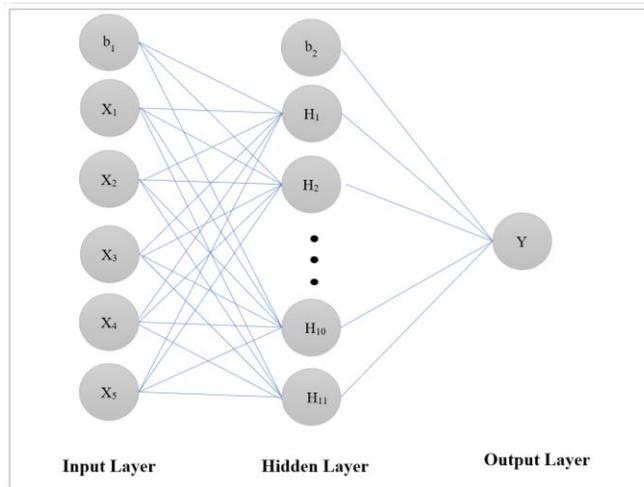
Fig. 6. ANN Architecture

Each one neuron in hidden layer learns the function $f(x)$ that can be seen in equation (2) [23] .

$$f(x) = W_2 g(W_1^T x + b_1) + b_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

where $g(z) = 1/(1 + e^{-z})$ which has a range of values $(0,1)$, W_1, W_2 consecutive represent the weights of the input layer and the weights of hidden layer. When the assign samples of outputs $f(x)$ value larger or equal 0.5, then it will be classified as rain class, and the rest to the not-rain class.

   Two main stages were carried out in this study to find predictive models, namely the model training stage and the testing stage. Fig. 7. show the training process using the backpropagation algorithm. The backpropagation algorithm is an example of an algorithm from supervised learning where the architecture consists of various interconnected layers. In addition, the backpropagation algorithm represents a type of Artificial Neural Network whose learning algorithm is based on the deepest-descent technique. If provided with an appropriate number of hidden units, they will be able to minimize errors from nonlinear functions with a high level of complexity. The training stage in ANN using backpropagation has 3 stages, namely feedforward, backward/backpropagation, and update weights and biases [22].
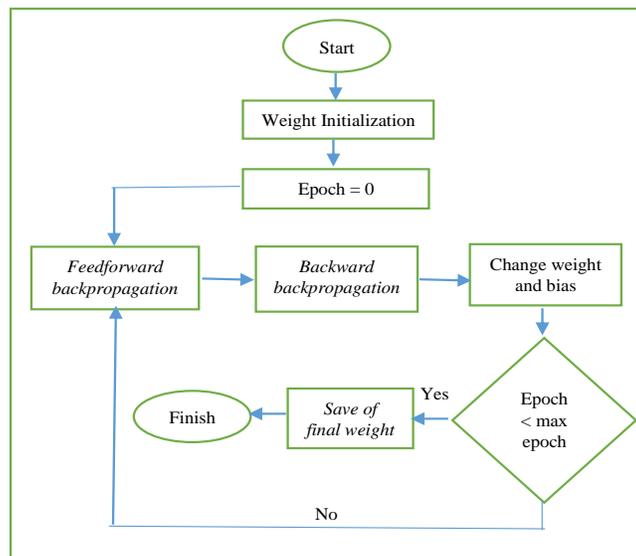

Fig. 7. The training process using the backpropagation algorithm

D. Testing

After performing 3 stages on the backpropagation algorithm, then the cessation condition test is then carried out, when it reaches the specified epoch then the final weight will be stored.

At the testing stage, the data to be used is a predetermined test data. At the testing stage, the last reading and taking of the weights stored in the previous training process will be carried out. Then, the feedforward or forward propagation stage is carried out again. After that, the accuracy calculation is carried out with the following formula:

$$Accuracy = \frac{\sum (Correct\ prediction)}{\sum (All\ predictions)} \tag{3}$$

In this study, to compare the effect of the use of different number of factors and the amount of data, analysis of variance (ANOVA) was used. Analysis of variance is a statistical method that can be used to detect differences between several experimental groups, with one or more independent variables. In ANOVA, the independent variables are called factors, while the groups within each factor are called levels. One of the advantages of ANOVA is its ability to analyze experimental designs consisting of several independent variables [24]. The hypothesis and analysis of the variance table are defined as follows:

H_0 : All scenarios give the same response

H_1 : At least there are scenario pairs that give different responses

IV. RESULTS AND DISCUSSION

In this study, two scenarios were carried out to obtain the best predictive model. The first scenario is based on the number of weather factors and the second scenario is based on the duration of the data used in the model training. The scenario based on the number of factors, used 5 weather factors including average temperature, rainfall, solar radiation, air pressure, and wind, and 4 weather factors including average temperature, solar radiation, air pressure, and wind. Meanwhile, based on the duration of the data used, the scenarios are 1 year, 2 years, 3 years to 6 years. The objectives to be achieved, related to the determination of the scenario used, to get the best model in predicting climate, to find out the amount of training data and the optimal number of factors in this climate prediction process.

The tests carried out aim to find out how effective the application of the Artificial Neural Network method with the backpropagation algorithm is on the same data as the two scenarios above. In building the prediction model, 4 factors and 5 factors are used as input layers, 11 hidden layers and 1 output layers (which produces 2 class labels). The number of neurons = 11 in hidden layers is set as the minimum (should be less than twice the size of the input layer.). The activation function used is a the logistic function $g(z) = 1/(1 + e^{-z})$ and the maximum epoch set is 50. The evaluation is carried out using the learning rate (LR) = 0.06.

While the scenario is based on the duration of the data used, at the time of testing 6 variations of the training data were used, namely:

TABLE IV
WEATHER FACTORS

No	Range	Duration (years)
1	2010 to 2015	6
2	2011 to 2015	5
3	2012 to 2015	4
4	2013 to 2015	3
5	2014 to 2015	2
6	2015 only	1

While the test data used is about 25% in the overall training data, namely data for 2016 and 2017. The entire training data is then used to build a neural network model using 4 factors and 5 weather factors.

The composition of the training data is intended to determine the effect of using the optimal amount of training data in making weather predictions. The results of testing using different amounts of training data can be seen in TABLE V.

TABLE V
ACCURACY RESULT OF USING AMOUNT OF TRAINING DATA FOR CLIMATE PREDICTION

Range of Training Data	Numbers of Factors Used	
	5 Factors	4 Factors
2015	58.33%	72.22%
2014 – 2015	61.11%	75%
2013 – 2015	77.78%	80.56%
2012 – 2015	77.78%	80.56%
2011 – 2015	80.56%	80.56%
2010 - 2015	83.33%	80.56%

Based on the experiment in TABLE V, it was found that the use of data on 4 weather factors for a period of 3 years or more as training data gave more stable weather prediction results, with an accuracy value of > 80%. The use of 5 weather factors to predict the monthly weather gives better results for large training data that is only 6 years. If the training data used is less than 6 years old, then the accuracy results obtained are not better than the model with 4 weather factors. This study is in accordance with [12], which investigated the prediction of average rainfall in Udupi Karnataka District with an artificial neural network model. The experimental results in this study show that the more the number of neurons in the ANN, the lower the Mean Square Error (MSE) and the greater the amount of input data, the lower the MSE after training, where input/output data normalization is performed if the interval is very high. Meanwhile [18], it was stated that the best MSE learning rate was 0.8. And with a maximum epoch of 50 iterations, a hidden layer consisting of 3 neurons and a learning rate of 0.6, obtained the lowest MSE {0.02815; 0.01686; 0.01934; 0.03196}.
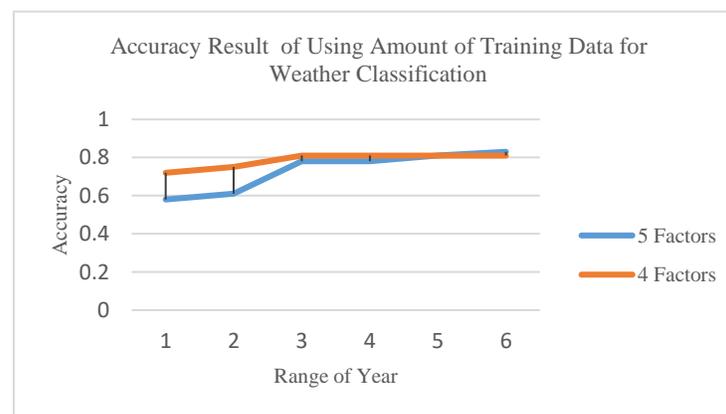

Fig. 8. Pattern of Accuracy Results 5 Factors Vs 4 Factors

While Fig. 8 describes the stability pattern of accuracy based on the number of factors and the amount of data. The X axis represents the amount of data described in the range of years, while the Y axis represents the accuracy value. Based on Fig. 8, it can be seen that the pattern of weather prediction accuracy results in the training data using 5 factors, which are indicated by the blue line, the accuracy will increase if the time span is

longer, in other words the amount of data is increasing. On the other hand, when using 4 factors, which are indicated by the red line, the accuracy pattern is more stable for a time span of more than 2 years. So that the more predictive accuracy factors for weather classification models with ANN, will be vulnerable to the amount of data, where the greater the amount of data, the accuracy will increase.

TABLE VI is a table that explains the significance of the effect of scenarios using the number of factors and the amount of data that varies, using analysis of variance (ANOVA). Based on TABLE VI, it can be seen that the use of a number of different factors in weather classification (rain or not) using ANN will have a significant effect on the 86% confidence interval, while the use of varying amounts of data or year ranges will have a significant effect on the 92% confidence interval.

TABLE VI
ANOVA OF TWO SCENARIOS

Source of variation	Sum Square	Degree of free	Mean Square	F Value	P Value	Critical F
Number of Data	507.656	5	101.531	4.009	0.077	5.05
Number of Factors	77.877	1	77.877	3.075	0.139	6.608
Error	126.619	5	25.324			
Total	712.153	11				

V. CONCLUSION

Weather classification in this study, was carried out using ANN, and a two-scenario approach, namely based on the number of factors and the amount of training data. The number of factors 4 and 5 will have a significant effect on the weather, with a confidence interval of 84%, while the use of the amount of training data will have a significant effect on the accuracy of the weather classification results, with a confidence interval of 92%. The experimental results show that the Neural Network model using 5 weather factors with 6 years of training data provides the highest accuracy, compared to the 4 factor model. However, if the training data used to form the model is less than or equal to 5 years, then the use of a model with 4 weather factors is more effective. Another output of this research is the production of training data modeling for the application of data mining classification techniques in the weather prediction process. This data model can be developed for the weather prediction process based on historical data from the previous few months.

REFERENCES

[1]   Julismin, "Dampak Dan Perubahan Iklim Di Indonesia," *J. Geogr.*, vol. 5, no. no.1, pp. 39–46, 2013.
[2]   O. C. P. Pradana and R. N. Sesanti, "ANALISIS DAMPAK PERUBAHAN IKLIM TERHADAP CURAH HUJAN BERDASARKAN PERUBAHAN ZONA AGROKLIMATOLOGI PADA SKALA LOKAL POLITEKNIK NEGERI LAMPUNG," *J. Wacana Pertan.*, vol. 14, no. 1, 2018, doi: http://dx.doi.org/10.37694/jwp.v14i1.23.
[3]   F. Sheikh, S. Karthick, D. Malathi, J. S. Sudarsan, and C. Arun, "Analysis of Data Mining Techniques for Weather Prediction," *Indian J. Sci. Technol.*, vol. 9, doi: 10.17485/ijst/2016/v9i38/101962.
[4]   D. Susilokarti, S. S. Arif, S. Susanto, and L. Sutiarso, "IDENTIFIKASI PERUBAHAN IKLIM BERDASARKAN DATA CURAH HUJAN DI WILAYAH SELATAN JATILUHUR KABUPATEN SUBANG, JAWA BARAT (Identification of Climate Change Based on Rainfall Data in Southern Part of Jatiluhur, Subang District, West Jawa)," *J. Agritech*, vol. 35, no. 01, p. 98, 2015, doi: 10.22146/agritech.13038.
[5]   S. D. Nandakumar, R. Valarmathi, P. Sudha Juliet, and G. Brindha, "Artificial Neural Network for Rainfall Analysis Using Deep Learning Techniques," *J. Phys. Conf. Ser.*, vol. 1964, no. 4, 2021, doi: 10.1088/1742-6596/1964/4/042022.
[6]   L. Chen and X. Lai, "Comparison between ARIMA and ANN models used in short-term wind speed

forecasting," *Asia-Pacific Power Energy Eng. Conf. APPEEC*, 2011, doi: 10.1109/APPEEC.2011.5748446.

[7]   M. Elsaraiti and A. Merabet, "A comparative analysis of the arima and lstm predictive models and their effectiveness for predicting wind speed," *Energies*, vol. 14, no. 20, 2021, doi: 10.3390/en14206782.

[8]   H. Jayadianti, T. A. Cahyadi, N. A. Amri, and M. F. Pitayandanu, "Metode Komparasi Artificial Neural Network Pada Prediksi Curah Hujan - Literature Review," *J. Tekno Insentif*, vol. 14, no. 2, pp. 48–53, 2020, doi: 10.36787/jti.v14i2.150.

[9]   N. Samsiahsani, I. Shlash, M. Hassan, A. Hadi, and M. Aliff, "Enhancing Malaysia Rainfall Prediction Using Classification Techniques Enhancing Malaysia Rainfall Prediction Using Classification Techniques," *J. Appl. Environ. Biol. Sci.*, vol. 7, no. 2s, pp. 20–29, 2017.

[10]   D. S. Jeslet and S. Jeevanandham, "Climate Change Analysis using Data Mining Techniques," *Int. J. Adv. Res. Sci. Eng.*, vol. 8354, no. 4, pp. 46–53, 2015.

[11]   R. Samya and R. Rathipriya, "Predictive Analysis for Weather Prediction using Data Mining with ANN : A Study," *Int. J. Comput. Intell. Informatics,* vol. 6, no. 2, pp. 149–153, 2016.

[12]   Q. Liu, "A survey on rainfall forecasting using artificial neural network Yanyun Zou Xiaodong Liu Nigel Linge," vol. 11, no. 2, pp. 240–249, 2019.

[13]   N. Thakur, S. Karmakar, and S. Soni, "Rainfall Forecasting Using Various Artificial Neural Network Techniques - A Review," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 3307, pp. 506–526, 2021, doi: 10.32628/cseit2173159.

[14]   F. Olaiya, "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies," *I.J. Inf. Eng. Electron. Bus.*, no. February, pp. 51–59, 2012, doi: 10.5815/ijieeb.2012.01.07.

[15]   ul M. Hasan, S. Ullah, M. J. Khan, and K. Khurshid, "Comparative analysis of SVM, ann and cnn for classifying vegetation species using hyperspectral thermal infrared data," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 42, no. 2/W13, pp. 1861–1868, 2019, doi: 10.5194/isprs-archives-XLII-2-W13-1861-2019.

[16]   J. Lee, C. G. Kim, J. E. Lee, N. W. Kim, and H. Kim, "Application of artificial neural networks to rainfall forecasting in the Geum River Basin, Korea," *Water (Switzerland)*, vol. 10, no. 10, 2018, doi: 10.3390/w10101448.

[17]   M. Akhter and M. A. Ahanger, "Climate modelling using ANN," *Int. J. Hydrol. Sci. Technol.*, vol. 9, no. 3, pp. 251–265, 2019, doi: 10.1504/IJHST.2019.102316.

[18]   A. N. Sihananto and W. F. Mahmudy, "Rainfall Forecasting Using Backpropagation Neural Network," *J. Inf. Technol. Comput. Sci.*, vol. 2, no. 2, pp. 66–76, 2017.

[19]   K. Abhishek, A. Kumar, R. Ranjan, and S. Kumar, "A rainfall prediction model using artificial neural network," *Proc. - 2012 IEEE Control Syst. Grad. Res. Colloquium, ICSGRC 2012*, no. July, pp. 82–87, 2012, doi: 10.1109/ICSGRC.2012.6287140.

[20]   A. Chaturvedi, "Rainfall Prediction using Back-Propagation Feed Forward Network," *Int. J. Comput. Appl.*, vol. 119, no. 4, pp. 1–5, 2015, doi: 10.5120/21052-3693.

[21]   N. Chamidah, W. Wiharto, and U. Salamah, "Pengaruh Normalisasi Data pada Jaringan Syaraf Tiruan Backpropagasi Gradient Pengaruh Normalisasi Data pada Jaringan Syaraf Tiruan Backpropagasi Gradient Descent Adaptive Gain (BPGDAG) untuk Klasifikasi," *J. ITSMART*, vol. 1 no 1, no. June, pp. 28–33, 2012, doi: 10.20961/itsmart.v1i1.582.

[22]   A. Alamsyah and M. F. Permana, "Artificial Neural Network for Predicting Indonesian Economic Growth using Macroeconomics Indicators," 2018.

[23]   "Neural network models." https://scikit-learn.org/stable/modules/neural_networks_supervised.html (accessed Dec. 29, 2021).

[24]   S. Sawyer, "Analysis of Variance : The Fundamental Concepts," no. December, 2017, doi: 10.1179/jmt.2009.17.2.27E.