Time Series Forecasting Covid 19 in Indonesia with a Perceptron Multilayer Neural Network

Reynaldi Thimotius Abineno[1]

[1]Program Studi Teknik Informatika, Universitas Kristen Wira Wacana Sumba

E-mail: reynaldi@unkriswina.ac.id,

KEYWORDS:

covid 19, artificial neural network, prediction.

ABSTRACT

Covid 19, which was first discovered in Wuhan, China, spread so rapidly throughout the world in early 2020 to date, according to data to date, the death rate due to Covid 19 has reached 1,338,100 people from a total of 55,624,562 positive cases worldwide. Prevention needs to be done to inhibit the rate of positive cases of Covid which are still increasing today. one of the efforts that can be done is to make predictions or forecasts. In this study, prediction or forecasting of positive cases of Covid will be carried out using time series data in the last few months using the backpropagation neural network method or multi-layer perceptron modeling to produce predictions with fairly good accuracy.

INTRODUCTION

Covid 19 is a virus that is dangerous because it attacks the respiratory tract, besides that the coronavirus is also dangerous because of its rapid transmission. Until now, it has been recorded that millions of people around the world have been infected with the disease.

One of the steps in anticipating a positive number of Covid 19 is to make predictions. prediction is done using an algorithm and based on time series data to produce a prediction for the future.

An artificial neural network is one of the models adopted from the human brain nerve cells. so the workings of this model are the same as the human brain. besides that, the artificial neural network is also quite good at making predictions. one of the models in neural networks is the backpropagation model or commonly known as the multi-layer perceptron where there are 3 layers in it, namely the input layer, hidden layer, and output layer. One of the advantages of this model is that it can minimize the error value due to the weight correction based on the error value.

In this study, the backpropagation model in Ann will be used to predict positive cases of Covid 19 in Indonesia.

RELATED WORKS

The artificial neural network method is one of the methods in data mining where this method can produce a prediction that can later be used as a parameter to control something accurately.

In a study conducted by Abdelhafid Zeroual et al (2020), they compared 5 algorithms used in predicting Covid 19 whose data were collected from 6 countries, namely Italy, Spain, France, China, the USA, and Australia. The 5 algorithms being compared are the simple Recurrent Neural Network (RNN) algorithm, Long short-term memory (L STM), Bidirectional L STM (BiL STM), Gated recurrent units (GRUs), and the Variational AutoEncoder (VAE) algorithm. The results of this study concluded that the Variational AutoEncoder (VAE) algorithm is superior in terms of accuracy compared to others. There are five parameters used, namely RMSE, MAE, MAPE, EV, RMSLE[1].

The research was conducted by Farhan Mohammad Khan (2020) to predict the number of Covid 19 in India based on time series data using the Auto-Regressive Integrated Moving Average (ARIMA) model. To compare the accuracy of the models used for prediction, a nonlinear autoregressive neural network (NAR) was developed. The highest value of an Arima model used is the value of the Bayesian Information Criteria (BIC) and the highest value of R^2 where the average result is 0.95% while the nonlinear autoregressive neural network (NAR) using 10 neurons produces an R^2 value of 0.97%[2].

In a study conducted by Xuan Zhang et al (2020), the backpropagation neural network method was combined with a genetic algorithm (GA) to predict dam seepage. the results will be compared with conventional statistical models which have a weakness because statistical models cannot show a non-linear relationship between dam seepage and load resulting in poor predictive results. Backpropagation neural networks and genetic algorithms (GA) increase the increase in nonlinear mapping and generalization so that the prediction results are more accurate shortly. so that the state of dam safety can be predicted[3].

Chen Zhang et al (2020). Conducted a study using an artificial neural network combined with a metaheuristic algorithm to predict the level of traffic accidents in Washington State for the period 2011-2015. The neural network will be carried out with 3 theoretical layers, then the Particle Swarm Optimization (PSO) method is used to improve the neural network. The result of combining backpropagation neural network and Particle Swarm Optimization (PSO) results in overall accuracy and speed. The results of the analysis also show that the factors related to vehicles have more influence on the number of accidents than those related to roads. The method applied in this research can be used in big data analysis and can be used in policymaking related to traffic accident rates[4].

in research conducted by Sourabh Shastri et al (2020). In this study, time series forecasting covid 19 was conducted using the LSTM-based RNN method. Short-term memory used is Convolution LSTM which can produce good accuracy. case studies are India and America[5].

in research conducted by Sina Ardabili et al (2020). Time series forecasting is done using the artificial neural network method with a system that uses the gray wolf optimization algorithm where predictions are made using time series data on January 22, 2020. The results of the model get good enough results to produce a prediction[6].

METHOD

This research will be conducted using a backpropagation neural network and covid 19-time series data sourced from the website. the research flow will be carried out as follows :

A. Rsearch workflow

The flow of this research will begin with a literature study where at this stage we will look for information about Covid 19 and also about the methods to be used. information search is done by looking for journals, books, or other sources that can be used as research references. After that the next stage is data search, the covid 19 data used is time-series data taken from the website. After that, enter the implementation stage where at this stage the backpropagation neural network method will be implemented to produce predictions. In the last stage, the implementation results will be analyzed to produce conclusions. the research flow will be carried out as follows :
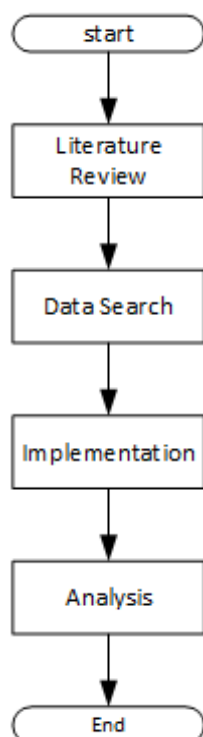
Figure 1. Research Workflow.

B. Artificial neural network (ANN)

Artificial neural networks are mathematical models or functions that resemble and represent the workings of the human brain. An artificial neural network consists of 3 layers, namely the input layer, hidden layer, and output layer. The input layer receives input data from the variable x then the hidden layer receives data from the input layer for processing and then passes it on to the output layer where the output layer calculates the variable x into a Y value.
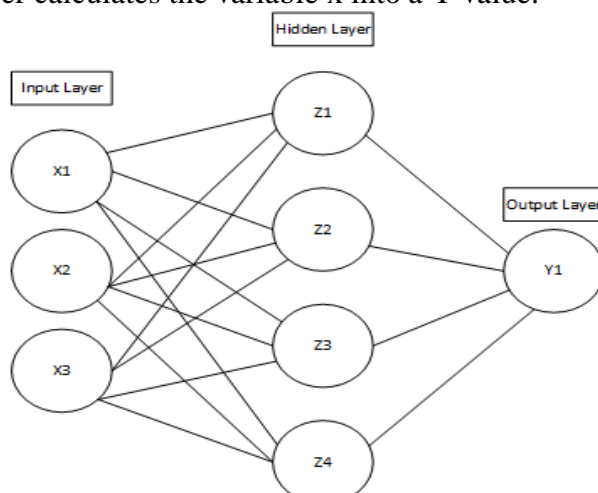

Figure 2. ANN architecture in general.

on each layer, there is a weight value that connects each layer. V_{ij} is the value that connects the input layer and the hidden layer. while $W_{j\square}$ is the weight value that connects the hidden layer to the output layer. other functions are as follows :

z_in_j is the result of data that has been processed in the hidden layer with the F1 activation function to produce Z_j.

$$z\_in_j = \sum_{i=1}^{n} x_i \cdot w_{ij}$$
$$z_j = F1(z_{in\,j})$$

y_in☐ is the result of processed data in the output layer with the activation function F2 to produce the Z_j value of the output layer.

$$y\_in_k = \sum_{j=1}^{n} z_i \cdot w_{jk}$$
$$z_j = F2\,(y_{in\,k})$$

C. Backpropagation

Backpropagation is algorithmic learning to minimize the error rate by adjusting the weights based on the difference between the output and the desired target. Backpropagation is also a systematic method for training multilayer neural networks. The hidden layer in backpropagation serves as a place to adjust the weight value until a new weight value is obtained that is appropriate and approaches the target. The activation function commonly used in backpropagation is the binary sigmoid activation function, where the output is 0 to 1.

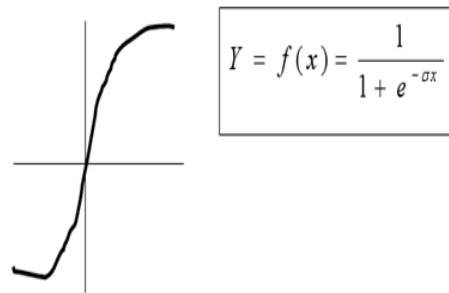$$Y = f(x) = \frac{1}{1 + e^{-\alpha x}}$$

Figure 3. the sigmoid binary activation function.

RESULT

In this study, predictions will be made using the time series data on positive cases of Covid 19 obtained through the https://www.kaggle.com/ page. The data used are data per April 1, 2020-18 November 2020. The graph of the addition of positive cases of Covid 19 in Indonesia can be seen in the following figure:
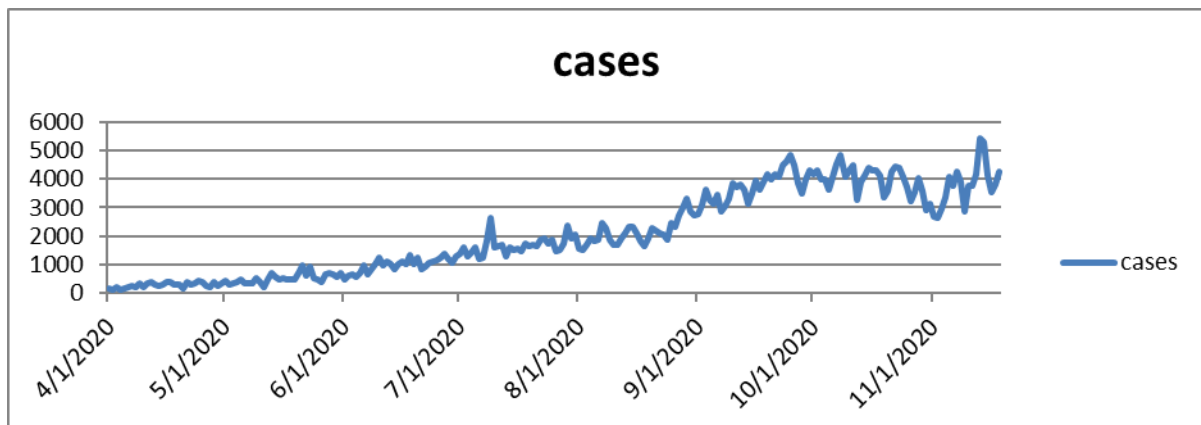

Figure 4. graph of additional cases of covid 19 in Indonesia

A. Workflow ann backpropagation

The stages and workflow of the artificial neural network used in this study are after the data set, the next step is preprocessing the data where at this stage we will normalize the data, then enter the network determination stage and also weighting. After initializing the network or weight, the feed-forward propagation process will be carried out and the process will be repeated with backpropagation which functions to adjust the weight value so that the error value can be minimized. Then from the trained data, predictions will be made. The prediction results will be denormalized to return to the original data form. workflow as shown below :
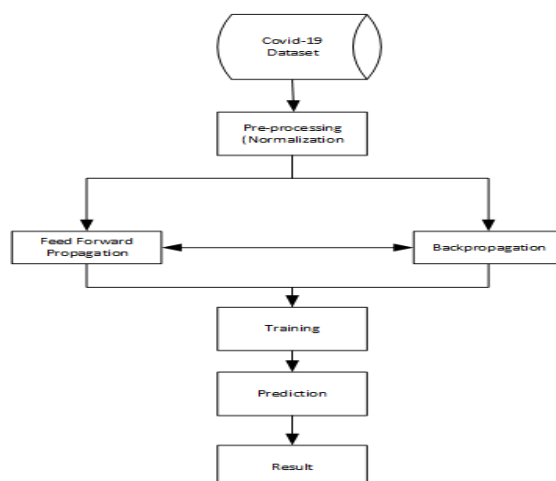

Figure 5. workflow ann backpropagation

B. Data normalization

Before predicting, the data that has been prepared must go through the data preprocessing stage, namely data normalization. Data normalization aims so that the output results can be processed according to the activation function used. The data that has been prepared will be normalized on a scale of 0-1. the equation used is the following equation:

$$X' = \frac{0,8\,(x-a)}{(b-a)} + 0,1$$

Normalized data is data on positive covid cases in Indonesia from April 1, 2020, to November 18, 2020. The following are the results of data normalization:

Table 1. normalization results

Date	Cases	X'
4/1/2020	149	0.106444
4/2/2020	113	0.101049
4/3/2020	196	0.113488
4/4/2020	106	0.1
4/5/2020	181	0.11124
4/6/2020	218	0.116785

4/7/2020	247	0.121132
4/8/2020	218	0.116785
4/9/2020	337	0.13462
4/10/2020	219	0.116935
4/11/2020	330	0.133571
4/12/2020	399	0.143912
4/13/2020	316	0.131472
4/14/2020	282	0.126377
4/15/2020	297	0.128625
4/16/2020	380	0.141064
4/17/2020	407	0.145111
4/18/2020	325	0.132821
4/19/2020	327	0.133121
4/20/2020	185	0.11184
4/21/2020	375	0.140315
4/22/2020	283	0.126527
4/23/2020	357	0.137617
4/24/2020	436	0.149457
4/25/2020	396	0.143462
4/26/2020	275	0.125328
4/27/2020	214	0.116186
4/28/2020	415	0.146309
4/29/2020	260	0.12308
4/30/2020	347	0.136118
…	…	…

C. Initialize network

The next stage is the network initialization stage where at this stage the appropriate number of hidden layers will be determined to produce predictive results that are closest to the actual data. so that we will use

2.2 hidden layers and 1 hidden input with a learning rate of 0.1 and epoch 1000. The network architecture is as follows:
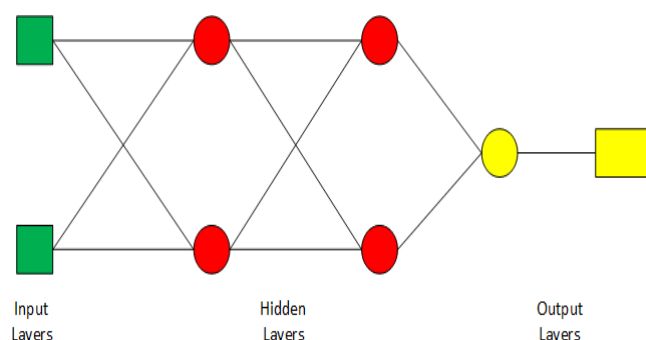

Figure 6. network architecture to be used

The network used is 2.2 hidden layers, and 1 output layer.

D. Training and prediction

after that, it will be practiced before making predictions. Training data will be carried out with data from 1 April 2020 to 18 November 2020, while the data used as test data are data from 19 November 2020 to 26 November 2020. predictions will be made using 2.2 hidden layers with a learning rate of 0.1 and epoch 1000.

predictions are made using the Weka machine learning application. where on weka The multilayer perceptron neural network method is equipped with the backpropagation learning method. Each neuron in one layer will be connected to other neurons in the next layer with a certain weight determined in the training process. in the backpropagation process, the weight value will affect the decrease in the resulting error value.

Result

The testing process can be done after the training process is complete. The results of predictions and forecasting will be denormalized first so that the predictive data returns to the data before normalization. The data denormalization process will be carried out with the following equation :

$$X = \frac{(x'-0,1)(b-a)}{0,8} + a$$

The prediction results are carried out from 19 November 2020 to 26 November 2020 and after denormalization, the results are as follows :

Table 2. prediction results

Date	actual data	prediction results
11/19/2020	4789	5263.843
11/20/2020	4792	5157.083

11/21/2020	4998	5079.014
11/22/2020	4360	5020.964
11/23/2020	4442	4976.925
11/24/2020	4192	4942.895
11/25/2020	5534	4916.873
11/26/2020	4917	4896.188

Table 2 shows the comparison between the actual data and the predicted results. where the architecture model used is 2.2 hidden layers and 1 output layer and the resulting results are almost close to the actual data. visualization of the prediction results can be seen in the following figure :
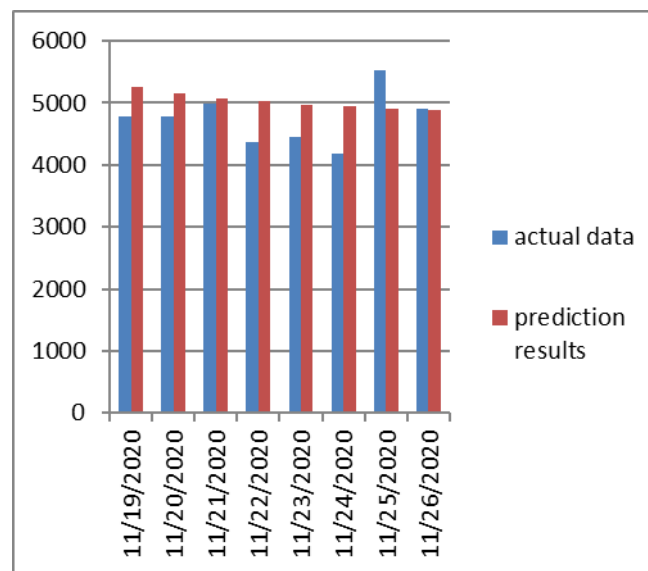
Figure 7. comparison of actual data and predicted results

In Figure 7 above, you can see the results of the comparison between the predicted results and actual data, where on 21 November 2020 and 26 November 2020 the predicted results have the smallest error or have the smallest difference with the actual data.

CONCLUSION AND FUTURE WORK

In this study, positive cases of Covid-19 in Indonesia are predicted using the multilayer perceptron backpropagation neural network method. testing was carried out using the Weka machine learning application. The model used is 2.2 hidden layers and 1 output layer with a learning rate of 0.1 and epoch

1000 capable of producing predictions that are close to the actual data. However, time-series forecasting in the weka itself must be carried out experimentally to get the most accurate results, where the data pattern of the number of positive cases of Covid 19 is still difficult to predict because the number of additions is sometimes too little and can be too much at the following times so that in further research it is necessary conducted experiments and tests with other models.

REFERENCES

[1]   A. Zeroual, F. Harrou, A. Dairi, and Y. Sun, "Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study," *Chaos, Solitons and Fractals*, vol. 140, 2020, doi: 10.1016/j.chaos.2020.110121.

[2]   F. M. Khan and R. Gupta, "ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India," *J. Saf. Sci. Resil.*, vol. 1, no. 1, pp. 12–18, 2020, doi: 10.1016/j.jnlssr.2020.06.007.

[3]   X. Zhang, X. Chen, and J. Li, "Improving Dam Seepage Prediction Using Back-Propagation Neural Network and Genetic Algorithm," *Math. Probl. Eng.*, vol. 2020, 2020, doi: 10.1155/2020/1404295.

[4]   C. Zhang *et al.*, "A Crash Severity Prediction Method Based on Improved Neural Network and Factor Analysis," *Discret. Dyn. Nat. Soc.*, vol. 2020, 2020, doi: 10.1155/2020/4013185.

[5]   S. Shastri, K. Singh, S. Kumar, P. Kour, and V. Mansotra, "Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study," *Chaos, Solitons and Fractals*, vol. 140, p. 110227, 2020, doi: 10.1016/j.chaos.2020.110227.

[6]   S. Ardabili, A. R. Varkonyi-Koczy, A. Mosavi, and S. S. Band, "Coronavirus Disease (COVID-19) Global Prediction Using Hybrid Artificial Intelligence Method of ANN Trained with Grey Wolf Optimizer," *medRxiv*, p. 2020.10.22.20217604, 2020, [Online]. Available: https://doi.org/10.1101/2020.10.22.20217604.

[7]   T. Chakraborty, A. Bhattacharyya, and M. Pattnaik, "Theta autoregressive neural network model for COVID-19 outbreak predictions," *medRxiv*, p. 2020.10.01.20205021, 2020, [Online]. Available: https://doi.org/10.1101/2020.10.01.20205021.

[8]   D. Mandal, S. K. Pal, and P. Saha, "Modeling of electrical discharge machining process using back propagation neural network and multi-objective optimization using non-dominating sorting genetic algorithm-II," *J. Mater. Process. Technol.*, vol. 186, no. 1–3, pp. 154–162, 2007, doi: 10.1016/j.jmatprotec.2006.12.030.

[9]   N. Hasan, "A Methodological Approach for Predicting COVID-19 Epidemic Using EEMD-ANN Hybrid Model," *Internet of Things*, vol. 11, p. 100228, 2020, doi: 10.1016/j.iot.2020.100228.

[10]   V. K. R. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks," *Chaos, Solitons and Fractals*, vol. 135, 2020, doi: 10.1016/j.chaos.2020.109864.

[11]   M. Maleki, M. R. Mahmoudi, M. H. Heydari, and K. H. Pho, "Modeling and forecasting the spread and death rate of coronavirus (COVID-19) in the world using time series models," *Chaos, Solitons and Fractals*, vol. 140, 2020, doi: 10.1016/j.chaos.2020.110151.

[12] Z. Ceylan, "Estimation of COVID-19 prevalence in Italy, Spain, and France," *Sci. Total Environ.*, vol. 729, p. 138817, 2020, doi: 10.1016/j.scitotenv.2020.138817.

[13] C. Katris, "A time series-based statistical approach for outbreak spread forecasting: Application of COVID-19 in Greece," *Expert Syst. Appl.*, vol. 166, no. April 2020, p. 114077, 2021, doi: 10.1016/j.eswa.2020.114077.

[14] A. H. Moghaddam, M. H. Moghaddam, and M. Esfandyari, "Predicción del índice del mercado bursátil utilizando una red neuronal artificial," *J. Econ. Financ. Adm. Sci.*, vol. 21, no. 41, pp. 89–93, 2016, doi: 10.1016/j.jefas.2016.07.002.

[15] O. A. Amalia and A. H. M. Putri, "Comparison of backpropagation artificial neural network and SARIMA in predicting the number of railway passengers," *J. Phys. Conf. Ser.*, vol. 1663, no. 1, 2020, doi: 10.1088/1742-6596/1663/1/012033.

[16] J. Gu, G. Yin, P. Huang, J. Guo, and L. Chen, "An improved back propagation neural network prediction model for subsurface drip irrigation system," *Comput. Electr. Eng.*, vol. 60, pp. 58–65, 2017, doi: 10.1016/j.compeleceng.2017.02.016.

[17] Z. Hu, Q. Ge, S. Li, L. Jin, and M. Xiong, "Artificial intelligence forecasting of covid-19 in China," *arXiv*, pp. 1–20, 2020.

[18] R. Olawoyin, "Application of backpropagation artificial neural network prediction model for the PAH bioremediation of polluted soil," *Chemosphere*, vol. 161, pp. 145–150, 2016, doi: 10.1016/j.chemosphere.2016.07.003.

[19] T. Chen, N. Kapron, and J. C. Y. Chen, "Using Evolving ANN-Based Algorithm Models for Accurate Meteorological Forecasting Applications in Vietnam," *Math. Probl. Eng.*, vol. 2020, 2020, doi: 10.1155/2020/8179652.

[20] B. Dai, H. Gu, Y. Zhu, S. Chen, and E. F. Rodriguez, "On the Use of an Improved Artificial Fish Swarm Algorithm-Backpropagation Neural Network for Predicting Dam Deformation Behavior," *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/5463893.