Analysis Accuracy of Artificial Neural Network Using Backpropagation Algorithm In Predicting Process (Forecasting)

Sandy Putra Siregar[1], Anjar Wanto[2]

[1]*STIKOM Tunas Bangsa Pematangsiantar*
[2]*STIKOM Tunas Bangsa Pematangsiantar*

[1]*sandyputra024@gmail.com,*[2]*anjarwanto@amiktunasbangsa.ac.id*

Abstract

Artificial Neural Networks are a computational paradigm formed based on the neural structure of intelligent organisms to gain better knowledge. Artificial neural networks are often used for various computing purposes. One of them is for prediction (forecasting) data. The type of artificial neural network that is often used for prediction is the artificial neural network backpropagation because the backpropagation algorithm is able to learn from previous data and recognize the data pattern. So from this pattern backpropagation able to analyze and predict what will happen in the future. In this study, the data to be predicted is Human Development Index data from 2011 to 2015. Data sourced from the Central Bureau of Statistics of North Sumatra. This research uses 5 architectural models: 3-8-1, 3-18-1, 3-28-1, 3-16-1 and 3-48-1. From the 5 models of this architecture, the best accuracy is obtained from the architectural model 3-48-1 with 100% accuracy rate, with the epoch of 5480 iterations and MSE 0.0006386600 with error level 0.001 to 0.05. Thus, backpropagation algorithm using 3-48-1 model is good enough when used for data prediction.

Keywords: *Analysis, Accuracy, ANN, Backpropagation, Prediction*

1. Introduction

1.1. Background

Prediction (forecasting) is basically a presumption about the occurrence of an event or event in the future. Prediction (forecasting) is very helpful in planning and decision-making activities of a policy [1]. Both government and private. Analysis of a prediction (forecasting) is very important to do in a study, so that research becomes more precise and directed [2]. Therefore required a good analysis by using methods that have been tested, so that the resulting accuracy can be really in accountability.

One method of prediction (forecasting) is widely used by researchers is to use artificial neural network backpropagation. Backpropagation algorithm is a method that can provide more accurate results in forecasting because this method performs repetitive training to get the best model and can also be analyzed in mathematics [3].

In this study the data to be used for prediction (forecasting) is the Human Development Index data. The Human Development Index (HDI) is a composite index that is influenced by health indicators represented by life expectancy, education indicators represented by the average length of school and economic indicators represented by people's purchasing power [4]. The Human Development Index (HDI) is structured as one of the alternative indicators, in addition to per capita national income, to assess development success undertaken by a country [5]. One of the reasons why human development needs to be addressed is: first, many developing countries including Indonesia have achieved economic growth, but have failed to reduce the social-economic and poverty gap [6].

Table 1. Human Development Index Data

No	Regency / City	HDI				
		2011	2012	2013	2014	2015
1	Asahan	65,87	66,23	66,58	67,51	68,40
2	Batu Bara	63,95	64,45	65,06	65,50	66,02
3	Binjai	70,85	71,54	72,02	72,55	73,81
4	Dairi	66,62	66,95	67,15	67,91	69,00
5	Deli Serdang	70,25	70,88	71,39	71,98	72,79
6	Gunungsitoli	63,71	64,34	65,25	65,91	66,41
7	Humbang Hasundutan	64,06	64,54	64,92	65,59	66,03
8	Karo	71,12	71,40	71,62	71,84	72,69
9	Labuhanbatu	67,88	68,64	69,45	70,06	70,23
10	Labuhanbatu Selatan	65,77	67,06	67,78	68,59	69,67
11	Labuhanbatu Utara	67,37	67,84	68,28	69,15	69,69
12	Langkat	65,77	66,18	67,17	68,00	68,53
13	Mandailing Natal	61,60	62,26	62,91	63,42	63,99
14	Medan	77,54	77,78	78,00	78,26	78,87
15	Nias	55,55	56,50	57,43	57,98	58,85
16	Nias Barat	55,43	56,20	56,58	57,54	58,25
17	Nias Selatan	55,50	55,97	56,78	57,78	58,74
18	Nias Utara	57,53	57,87	58,29	59,18	59,88
19	Padang Lawas	63,28	64,05	64,62	65,50	65,99
20	Padang Lawas Utara	65,22	65,65	66,13	66,50	67,35
21	Padangsidimpuan	71,08	71,38	71,68	71,88	72,80
22	PakpakBharat	63,11	63,88	64,73	65,06	65,53
23	Pematangsiantar	73,61	74,51	75,05	75,83	76,34
24	Samosir	65,81	66,31	66,80	67,80	68,43
25	Serdang Bedagai	65,28	66,14	67,11	67,78	68,01
26	Sibolga	69,17	69,71	70,45	71,01	71,64
27	Simalungun	69,03	69,79	70,28	70,89	71,24
28	Tanjungbalai	64,13	64,89	65,40	66,05	66,74
29	Tapanuli Selatan	65,14	65,95	66,75	67,22	67,63
30	Tapanuli Tengah	65,16	65,43	65,64	66,16	67,06
31	Tapanuli Utara	69,24	69,83	70,50	70,70	71,32
32	Tebing Tinggi	70,84	71,34	71,85	72,13	72,81
33	Toba Samosir	71,39	71,89	72,36	72,79	73,40

Source: Central Bureau of Statistics of North Sumatra

1.2. Formulation of the problem

The formulation of the problem of this study is How the accuracy of the prediction index of human development by using backpropagation algorithm.

1.3. Scope of problem

This study only predicts the index of human development in the regency/city in North Sumatra province and will only conduct a discussion of accuracy that will result from backpropagation algorithm against human development index prediction.

1.4. Research purposes

The purpose of this research is to conduct a learning/training process on the backpropagation algorithm, so that later it will produce an accurate and practical prediction, especially on human development index.

1.5. Benefits of research

The authors hope later this research is useful for other researchers who are doing research in the field of artificial neural networks specifically related to the human development index and beneficial to the government in economic policymaking.

2. Rudimentary

2.1. Analysis of ANN model

The schematic representation of artificial neural network. It consists of: (1) input layer (independent variables), (2) hidden layers and, (3) the output layer (dependent variable) [7].

2.2. Accuracy

The generalization accuracy is optimized over the empirical error and the flatness of the regression. On the other hand, ANN models are based on the empirical risk minimization (ERM) principle which only minimizes the empirical error and does not consider the capacity of the learning machines [7].

2.3. Artificial Neural Networks (ANN)

Artificial Neural Networks (ANNs) are intelligent methods in advanced computing that quantitatively analyze information by learning and training, the same way as human intelligent systems [8]. Artificial neural network (ANN) is one of the methods that is suitable to deal with the internal relations of complex model because of its highly nonlinear, large amounts of data parallel processing, high robustness, and fault tolerance [9]. Artificial Neural Network (ANN) is a computational model, which is based on Biological Neural Network. Artificial Neural Network is often called as Neural Network (NN) [10].

2.4. Backpropagation

Backpropagation (BP) algorithm was used to develop the ANN model [11]. The typical topology of BPANN (Backpropagation Artificial Neural Network) involves three layers: input layer,where the data are introduced to the network; hidden layer, where the data are processed; and output layer,where the results of the given input are produced [12]. A back-propagation algorithm was used for training. It is a convenient and simple iterative algorithm that usually performs well, even with complex data. Unlike other learning algorithms (like Bayesian learning) it has good computational properties when dealing with largescale data [13]. Backpropagation training method involves feedforward of the input training pattern, calculation and backpropagation of error, and adjustment of the weights in synapses [14].

2.5. Prediction (Forecasting)

Prediction or Forecasting is applications-oriented. It illustrates all methods with detailed real-world applications that reect typical forecasting situations. In various places, it uses applications not simply to illustrate the methods but also to drive home an important lesson via truly realistic examples: not everything works perfectly in the real world. Forecasting is in touch with modern modeling and forecasting software [15].

3. Research Methods

The research methodology is the stages in conducting a research to collect data or information that will be used in finding a solution of the problems that have been found.

3.1. Research Framework

In this research will be explained the flow of methodology or research framework. The framework is indispensable in conducting a study. The research framework starts from Data Collection; Data collection is done from Badan Pusat Statistik in 2011 until 2015. Then proceed with Library Studies; In addition to data retrieval, literature study one way to find or complement the research being done as a reference to make the research becomes more complete. Then Identify the Problem; Identifying the problem is one of the most necessary things with the data being met and converted as required and the specified weight. Then Preprocess; At this stage data identification or changes that make the data more consistent, and eliminate redundant data. After that Determine the Model; Model used in research of artificial neural network backpropagation to do prediction. After completion of the data processing phase and the model, a determination is completed the data test is done using software Matlab. Then do the prediction; Predictions are made for Comparison of the number of Artificial Neural Networks with the most accurate Backpropagation method. The final stage is to evaluate whether the test data processing as expected.
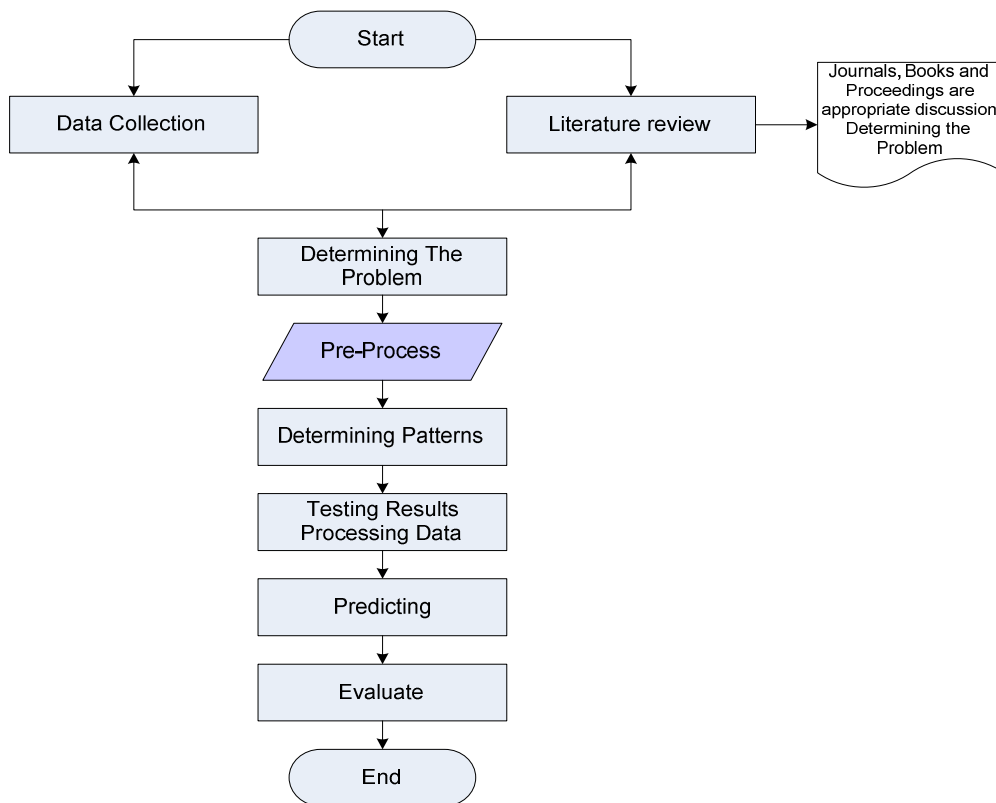

Figure 1. Research Framework

3.3. Data Retrieval

Data collection Human Development Index (HDI) by Regency / City comes from North Sumatra Central SATTATOR Agency 2011 to 2015 (Can be seen in table 1).

3.3. Data Normalization

Before it is processed, the data is normalized first by using the Sigmoid function, represented by equation (1).

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

Based on table 1, datasets from 2011 to 2013 are used as Training data, while datasets from 2012 to 2014 are used as data testing. In training data, the maximum value (b) of the dataset is 78.26. While the minimum value (a) is 55.43. By using the sigmoid function it will get the normalization data as follows:

$$x' = \frac{0.8(65,87 - 55,43)}{78,26 - 55,43} + 0.1$$

Then will be obtained results Normalization data 1 the year 2011 is 0.47 and so on. As for Testing data, The maximum value (b) of the dataset is 78.87. The minimum value (a) is 55.97. By using the same formula, it will get data 1, 2012 is 0.46, and so on until completion of normalization.

Table 2. Normalization of Training Data

Data	Input (In Years)			Target
	2011	2012	2013	
1	0,47	0,48	0,49	0,52
2	0,40	0,42	0,44	0,45
3	0,64	0,66	0,68	0,70
4	0,49	0,50	0,51	0,54
5	0,62	0,64	0,66	0,68
6	0,39	0,41	0,44	0,47
7	0,40	0,42	0,43	0,46
8	0,65	0,66	0,67	0,68
9	0,54	0,56	0,59	0,61
10	0,46	0,51	0,53	0,56
11	0,52	0,53	0,55	0,58
12	0,46	0,48	0,51	0,54
13	0,32	0,34	0,36	0,38
14	0,87	0,88	0,89	0,90
15	0,10	0,14	0,17	0,19
16	0,10	0,13	0,14	0,17
17	0,10	0,12	0,15	0,18
18	0,17	0,19	0,20	0,23
19	0,38	0,40	0,42	0,45
20	0,44	0,46	0,47	0,49
21	0,65	0,66	0,67	0,68
22	0,37	0,40	0,43	0,44
23	0,74	0,77	0,79	0,81
24	0,46	0,48	0,50	0,53
25	0,45	0,48	0,51	0,53
26	0,58	0,60	0,63	0,65
27	0,58	0,60	0,62	0,64
28	0,40	0,43	0,45	0,47
29	0,44	0,47	0,50	0,51
30	0,44	0,45	0,46	0,48
31	0,58	0,60	0,63	0,64
32	0,64	0,66	0,68	0,69
33	0,66	0,68	0,69	0,71

Table 3. Normalization of Testing Data

Data	Input (In Years)			Target
	2012	2013	2014	
1	0,46	0,47	0,50	0,53
2	0,40	0,42	0,43	0,45
3	0,64	0,66	0,68	0,72
4	0,48	0,49	0,52	0,56
5	0,62	0,64	0,66	0,69
6	0,39	0,42	0,45	0,46
7	0,40	0,41	0,44	0,45
8	0,64	0,65	0,65	0,68
9	0,54	0,57	0,59	0,60
10	0,49	0,51	0,54	0,58
11	0,51	0,53	0,56	0,58
12	0,46	0,49	0,52	0,54
13	0,32	0,34	0,36	0,38
14	0,86	0,87	0,88	0,90
15	0,12	0,15	0,17	0,20
16	0,11	0,12	0,15	0,18
17	0,10	0,13	0,16	0,20
18	0,17	0,18	0,21	0,24
19	0,38	0,40	0,43	0,45
20	0,44	0,45	0,47	0,50
21	0,64	0,65	0,66	0,69
22	0,38	0,41	0,42	0,43
23	0,75	0,77	0,79	0,81
24	0,46	0,48	0,51	0,54
25	0,46	0,49	0,51	0,52
26	0,58	0,61	0,63	0,65
27	0,58	0,60	0,62	0,63
28	0,41	0,43	0,45	0,48
29	0,45	0,48	0,49	0,51
30	0,43	0,44	0,46	0,49
31	0,58	0,61	0,61	0,64
32	0,64	0,65	0,66	0,69
33	0,66	0,67	0,69	0,71

4. Analysis And Results

4.1. Analysis

The first thing to do to program backpropagation with Matlab software is to initialize the network.

Determine the desired parameter value before doing the Training in order to obtain optimal results. After the parameters have been determined, then forming the data table in such a way as to be easy in processing and does not occur inconsistent data and avoid redundant data.

Table 4. Architecture ANN Backpropagation

Characteristics	Specification
Architecture	1 hidden layer
Input neuron	3
Hidden Neuron	8, 18, 28, 16, 48
Activation Function	Sigmoid
Goal	0.001
Initialization weights	Random
Minimum Error	0.001 – 0.05
Maximum Epoch	10000
Learning Rate	0.01

4.2. Results

This research uses 5 architecture that is 3-8-1, 3-18-1, 3-28-1, 3-16-1, 3-48-1. Of the 5 best architectural architecture is 3-48-1 with an accuracy of 100% and epoch as much as 5480 iterations.
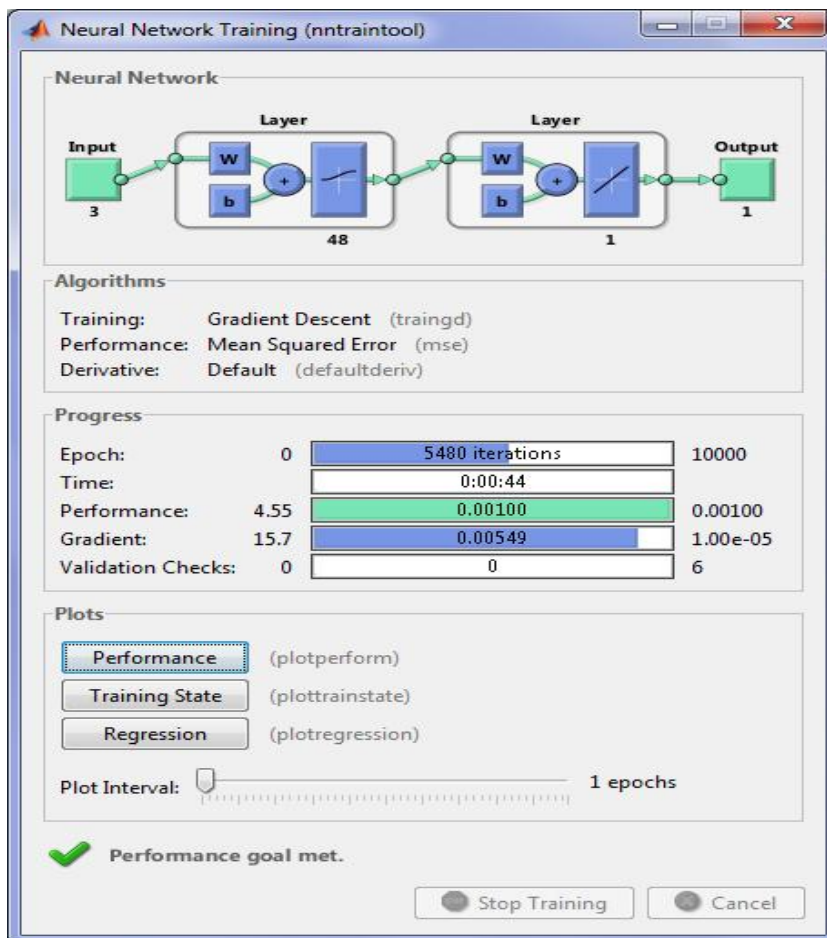

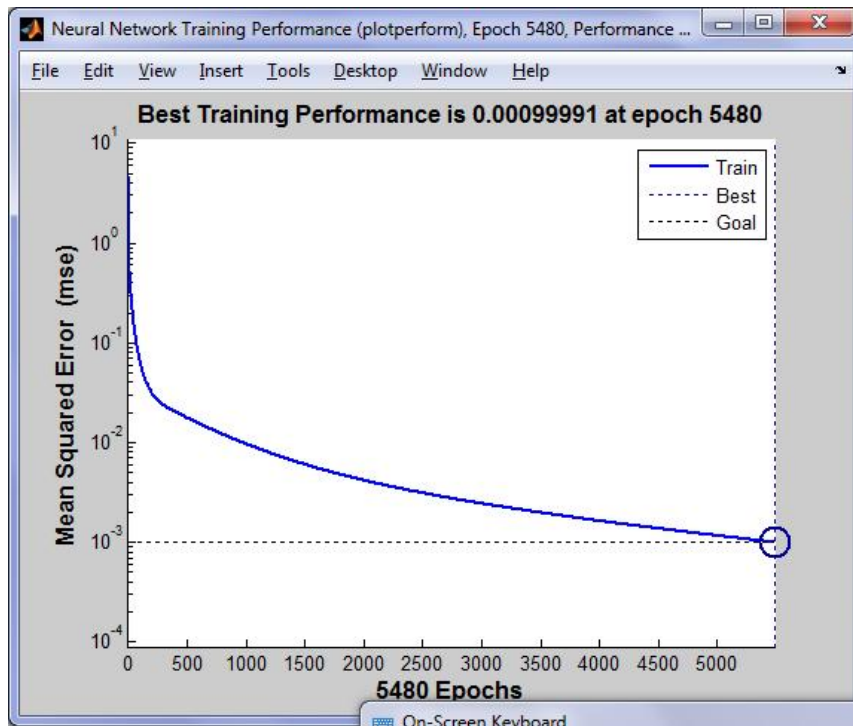**Figure 2. Training Results Model 3-48-1**

Figure 3. Results of Performance Training Model 3-48-1

Comparison of predicted results and accuracy rates using backpropagation algorithms vary greatly. Based on table 5, there are two architectural models whose accuracy is 100%. But the authors prefer 3-48-1 models, because they have epoch, time and MSE testing which is smaller than the 3-16-1 model.

Table 5. Data Training And Testing Backpropagation

No	Architecture	Training			Testing	
		Epoch	Time	MSE	MSE	Accuracy
1	3-8-1	1709	00:13	0,0009423	0,0012270	91%
2	3-18-1	3361	00:25	0,0148224	0,0007664	97%
3	3-28-1	8195	01:00	0,0009956	0,0017550	82%
4	3-16-1	9757	01:10	0,0010494	0,0006968	100%
5	3-48-1	5480	00:40	0,0010627	0,0006387	100%

5. Conclusion And Suggestion

Based on the description of the results of the previous discussion can be concluded that :

a. Architecture Model 3-48-1 can predict Human Development Index with 100% accuracy.
b. Architectural models greatly affect the accuracy of Training and Testing.
c. Variations will occur when testing data with 5 different architectural models.

References

[1] R. Sri *et al.*, "Application of Singular Spectrum Analysis (SSA) Methods in Indonesia Economic Growth Forecasting in 2017," *Journal Mathematics "MANTIK,"* vol. 3, no. 1, pp. 5–12, 2017.

[2] A. Wanto and A. P. Windarto, "Analisis Prediksi Indeks Harga Konsumen Berdasarkan Kelompok Kesehatan Dengan Menggunakan Metode Backpropagation," *Sinkron Jurnal & Penelitian Teknik Informatika*, vol. 2, no. 2, pp. 37–44, 2017.

[3] A. Mulyani, "Neural Network Structure Backpropagation Analysis For Forecasting Method In Calculating Poverty Level In Indonesia," *Journal Techno Nusa Mandiri*, vol. XIII, no. 1, pp. 9–15, 2016.

[4] I. D. Syahputra and A. Karim, "Comparison of Ordinary Least Square (OLS) Methods And Robust Regression," *National Seminar on Education, Science and Technology Faculty of Mathematics and Natural Sciences Muhammadiyah University of Semarang*, 2015, vol. 3, no. 1, pp. 18–31.

[5] R. U. Harahap, "The Effect of General Allocation Funds, Special Allocation Funds, And Profit Sharing Funds On Human Development Index In Regency / City Of Province Of North Sumatra," *KITABAH - Journal of Accounting and Sharia Finance*, vol. 1, pp. 118–130, 2017.

[6] M. Pangestika and E. Widodo, "Panel Regression Analysis Of Factors Affecting Human Development Index In Districts / Cities D.I.Yogyakarta," *National Seminar and The 4th Call for Sharia Paper*, 2017, pp. 198–205.

[7] N. Parveen, S. Zaidi, and M. Danish, "Development of SVR-based model and comparative analysis with MLR and ANN models for predicting the sorption capacity of Cr(VI)," *Process Safety Environmental Protection.*, vol. 107, no. vi, pp. 428–437, 2017.

[8] R. Olawoyin, "Application Of Backpropagation Artificial Neural Network Prediction Model For The PAH Bioremediation Of Polluted Soil," *Journal Chemosph.*, vol. 161, pp. 145–150, 2016.

[9] Z.-H. Wang, D.-Y. Gong, X. Li, G.-T. Li, and D.-H. Zhang, "Prediction Of Bending Force In The Hot Strip Rolling Process Using Artificial Neural Network And Genetic Algorithm (ANN-GA)," *Int. J. Adv. Manuf. Technol.*, 2017.

[10] Sumijan, A. P. Windarto, A. Muhammad, and Budiharjo, "Implementation of Neural Networks in Predicting the Understanding Level of Students Subject," *International Journal Of Software Engginering and Its Application*, vol. 10, no. 10, pp. 189–204, 2016.

[11] P. Antwi *et al.*, "Estimation of Biogas and Methane Yields in an UASB Treating Potato Starch Processing Wastewater With Backpropagation Artificial Neural Network," *Journal Bioresour Technology*, vol. 228, pp. 106–115, 2017.

[12] D. Huang and Z. Wu, "Forecasting Outpatient Visits Using Empirical Mode Decomposition Coupled With Backpropagation Artificial Neural Networks Optimized by Particle Swarm Optimization," *PLoS One*, vol. 12, no. 2, pp. 1–18, 2017.

[13] A. Ehret, D. Hochstuhl, D. Gianola, and G. Thaller, "Application of neural networks with back-propagation to genome-enabled prediction of complex traits in Holstein-Friesian and German Fleckvieh cattle," *Genetics Selection Evolution*, vol. 47, no. 1, p. 22, 2015..

[14] J. Tarigan, Nadia, R. Diedan, and Y. Suryana, "Plate Recognition Using Backpropagation Neural Network and Genetic Algorithm," *Procedia Computer Science*, vol. 116, pp. 365–372, 2017.

[15] F. X. Diebold, *Forecasting in Economics, Business, Finance and Beyond*. 2017.

Authors

1st Author
Sandy Putra Siregar
College Student of STIKOM Tunas Bangsa Pematangsiantar
sandyputra024@gmail.com

2nd Author
Anjar Wanto
Lecturer of STIKOM Tunas Bangsa Pematangsiantar
anjarwanto@amiktunasbangsa.ac.id