International Journal of Information System & Technology
Vol. 3, No. 1, (2019), pp.98-106

Architectural Model of Backpropagation ANN for Prediction of Population-Based on Sub-Districts in Pematangsiantar City

Marseba Situmorang[1], Anjar Wanto[2], Zulaini Masruro Nasution[3]

[1,2,3] STIKOM Tunas Bangsa Pematangsiantar
[1] marsebasitumorang465@gmail.com, [2] anjarwanto@amiktunasbangsa.ac.id,
[3] zulaini@amiktunasbangsa.ac.id

Abstract

A population is a group of individuals who occupy or live in a place or area that interacts with one another. Because the population has a very important role in an area, it is important to make predictions to find out how much the level of increase or descent of the population in an area, especially in Pematangsiantar. Therefore this research was conducted. This study uses population data in 8 Sub-Districts in Pematangsiantar. Data was taken from the Central Statistics Agency (BPS) of Pematangsiantar city in 2011-2017. The method used is the Artificial Neural Network (ANN) Backpropagation. These data will be processed into 2 parts namely training data and Testing data. This research will use 5 architectural models namely, 3-25-1, 3-30-1, 3-45-1, 3-54-1 and 3-68-1. From these 5 architectural models, after analysis, models 3-45-1 were chosen as the best models with epoch 553 values, MSE training 0,0001108768, MSE testing 0.0012355953 and an accuracy rate of 88%. The results of this paper are expected to be widely useful, especially for academics as further research material, especially those related to population in Pematangsiantar, because this research is still limited to discussing the level of accuracy, not prediction results.

Keywords: *Architectural Model, ANN, Population, Sub-Districts, Pematangsiantar*

1. Introduction

Residents are all people who have lived or are domiciled in the geographical area of the Republic of Indonesia for 6 months or more and or who are domiciled for less than 6 months but aim to settle. Total population means the number of residents who live/settle in a certain area/region, which is measured in units of souls per year. The influence of the population at a moderate level is basically positive and beneficial for economic development, both for developed and developing countries [1]. In Pematangsiantar itself, the population is quite dense, because Pematangsiantar is the second-largest city in North Sumatra after Medan. This research is expected to be widely used, especially for academics as research material specifically related to population, so the results can indirectly help the Pematangsiantar regional government in analyzing population growth and growth in the region.

The problem of population is an important problem in developing an area, decreasing or increasing population in an area has a very important role in the region itself. As we know that almost all development plans need to be supported with data on population, distribution, and composition to be relevant to the plan, not only on development plans that require population data. But in terms of the economy, education, health and so on. Then the complicated problems for the government in an effort to build and improve the standard of living of the community increases, if the population in an area is higher [2]. Pematangsiantar has 8 Sub-Districts whose population is quite dense, including Siantar Marihat, Siantar Marimbun, South Siantar, West Siantar, North Siantar, East Siantar, Siantar Martoba, and Siantar Sitalasari. This, of course, will affect development in the Sub-District. Therefore it is necessary to make predictions on the population in each Sub-

District so that the local government can anticipate excessive population growth in the context of development and improvement of community welfare in Pematangsiantar City.

The discussion in this paper focuses on analyzing the accuracy that results from the application of backpropagation neural network algorithms based on population in 9 Sub-Districts in Pematangsiantar. The reason for implementing this backpropagation artificial neural network is because this method is able to make predictions based on past data that has already taken place [3]–[10]. Previously there had been studies that discussed population growth predictions, namely the number of residents in Simalungun district using the Backpropagation algorithm. This research resulted in an accuracy rate of 97%, with an error rate of 0.01-0001, bipolar sigmoid (tansig) activation function, gradient descent (traingd) training function and the best architectural model using 2 hidden layers (5-10) namely 3-5 -10-1 [11]. Therefore, the authors conducted this research using the same algorithm, Backpropagation, with the same error rate parameters 0.01-0001 but by using a different hidden layer, namely 1 hidden layer only, the function of binary sigmoid activation (logsig) and functions Gradient descent training with momentum and adaptive learning rate backpropagation (traingdx).

2. Research Methodology

2.1. Data Source

This study uses population data in 8 Sub-Districts in Pematangsiantar. Data was taken from the Central Statistics Agency (BPS) of the city of Pematangsiantar in 2011-2017.

Table 1. Population Data by Sub-District in Pematangsiantar

No	Sub-District	Years				
		2013	2014	2015	2016	2017
1	Siantar Marihat	18196	18191	18274	18867	19096
2	Siantar Marimbun	14905	14884	14946	15427	15607
3	Siantar Selatan	17151	17150	17169	17726	17859
4	Siantar Barat	35438	35467	35587	36731	37125
5	Siantar Utara	46608	46613	46659	48165	48359
6	Siantar Timur	38570	38613	38646	39893	40202
7	Siantar Martoba	38759	38750	38831	40086	40466
8	Siantar Sitalasari	27266	27279	27322	28209	28517
Pematangsiantar		236893	236947	237434	245104	247411

Source: The Central Statistics Agency (BPS) of Pematangsiantar City

2.2. Backpropagation

The Backpropagation learning algorithm (BPLA)has become famous learning algorithms among ANNs. Backpropagation ANNs have been widely and successfully applied in diverse applications, such as patternrecognition, location selection and performance evaluations [12]. BPANN is the most extensively used ANN model. The typical topology of BPANN 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 [13]. Backpropagation training method involves feedforward of the input training pattern, calculation and backpropagation of error, and adjustment of the weights in synapses [14].

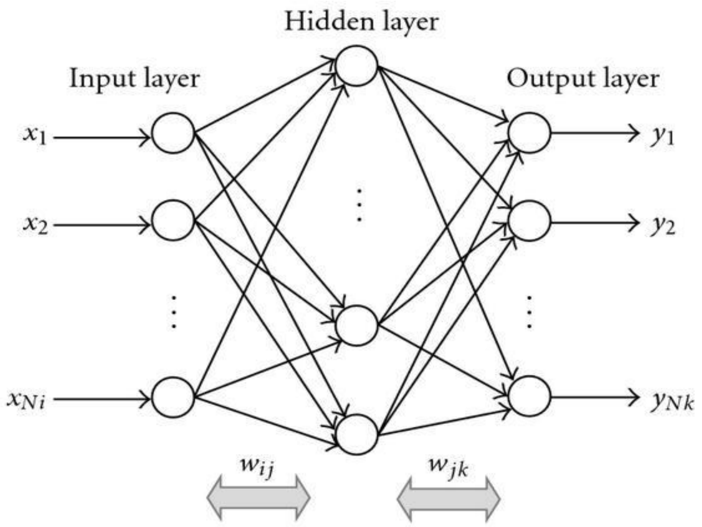

Figure 1. Schematic diagram of a general back-propagation neural network

2.3. Research Flow

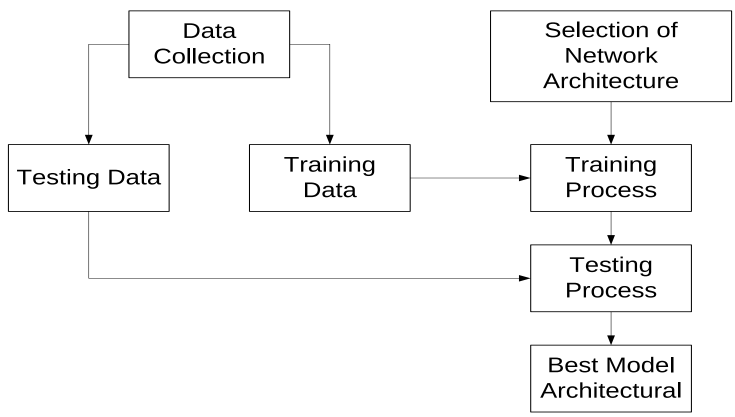

Figure 2. Research Flow

In Figure 2 it can be explained that the first thing to do is collect the dataset. The dataset used is data on the number of residents of 8 subdistricts in Pematangsiantar. Then preprocessing and dividing the data is done into several parts, namely the data used for training and the data used for testing. Then determine the network architecture model that will be used for the training process and the testing process, after all, that is done will be obtained based on the architectural model used. Next, from the several architectural models used, each was analyzed and then the best was chosen.

2.4. Normalization Formula

The data in table 1 will be normalized using the following formula [15]–[26]:

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

Explanation: x' is the result of normalization, x is data that will be normalized, a is the lowest data and b is the highest data from the dataset.

3. Results and Discussion

3.1. Normalized Result Data

The research data in table 1 will be divided into 2 parts, namely training data and testing data. The training data used is the 2011-2013 data with the 2014 target. While the test data is taken from 2014-2016 with the 2017 target. For the Normalization results of training, data can be seen in table 2 below.

Table 2. Normalization of Training Data

No	Sub-District	2011	2012	2013	Target (2014)
1	Siantar Marihat	0,11151	0,11149	0,11178	0,11384
2	Siantar Marimbun	0,10007	0,10000	0,10022	0,10189
3	Siantar Selatan	0,10788	0,10787	0,10794	0,10988
4	Siantar Barat	0,17142	0,17152	0,17194	0,17592
5	Siantar Utara	0,21024	0,21026	0,21042	0,21565
6	Siantar Timur	0,18231	0,18246	0,18257	0,18690
7	Siantar Martoba	0,18296	0,18293	0,18321	0,18758
8	Siantar Sitalasari	0,14303	0,14307	0,14322	0,14630

As for the results of Normalization of test data can be seen in table 3 below.

Table 3. Normalization of Testing Data

No	Sub-District	2014	2015	2016	Target (2017)
1	Siantar Marihat	0,11166	0,11243	0,11298	0,11364
2	Siantar Marimbun	0,10000	0,10061	0,10106	0,10159
3	Siantar Selatan	0,10779	0,10824	0,10875	0,10914
4	Siantar Barat	0,17219	0,17352	0,17459	0,17576
5	Siantar Utara	0,21093	0,21159	0,21359	0,21470
6	Siantar Timur	0,18290	0,18395	0,18510	0,18602
7	Siantar Martoba	0,18356	0,18485	0,18601	0,18714
8	Siantar Sitalasari	0,14331	0,14436	0,14518	0,14609

In this study, data processing was assisted with the 2011b Matlab tool in determining the best architectural model with Backpropagation. The architecture is used as many as 5 models, namely: namely 3-25-1, 3-30-1, 3-45-1, 3-54-1 and 3-68-1. How to determine the best architectural model with the Backpropagation method is to determine the minimum error of the training and testing process carried out. The error rate used is 0.01-0001. In this study, the code parameters used were analyzed using the Matlab 2011b application which can be seen in the following table 4.

Table 4. Program Parameters and Codes

Training Code	*Testing Code*
>> net=newff(minmax(P),[hidden layer,outputlayer],{'logsig','purelin'},'traingdx'); *>> net.IW{1,1};* *>> net.b{1};* *>> net.LW{2,1};* *>> net.b{2};* *>> net.trainParam.epochs=10000;* *>> net.trainParam.goal = 0,001;* *>> net.trainParam.lr=0.01;* *>> net.trainParam.show = 1000;* *>> net=train(net,P,T)* *[a,Pf,Af,e,Perf]=sim(net,P,[],[],T)*	*>> PP=[input data pengujian]* *>> TT=[output pengujian]* *[a,Pf,Af,e,Perf]=sim(net,PP,[],[],TT)*

3.2. Training and Testing

There are 5 training results and 5 testing results using each architectural model, namely 3-25-1, 3-30-1, 3-45-1, 3-54-1 and 3-68-1. However, the author will only write and explain the best architectural model, 3-45-1. For the results of training using architectural models, 3-45-1 can be seen in the following figure.

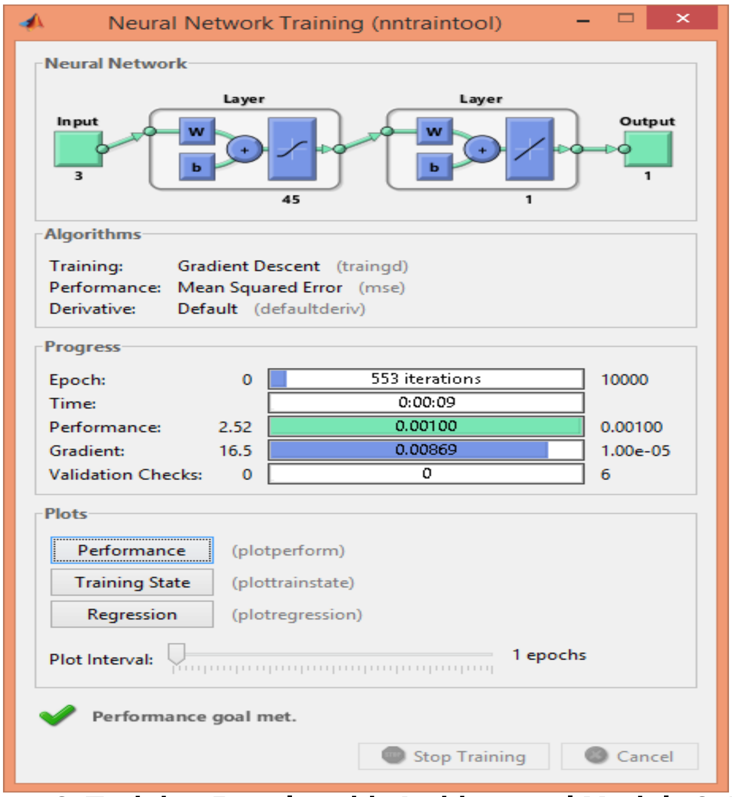

Figure 3. Training Results with Architectural Models 3-45-1

Based on figure 3 it can be explained that the results of the training using models 3-45-1 produce an epoch of 553 iterations with 9 seconds, and this model is the best architecture compared to the other 4 models. For training and testing tables can be seen in table 5 and table 6 below.

Table 5. Training with Models 3-45-1

No	Sub-District	Target	Output	Error	SSE
1	Siantar Marihat	0,11384	0,11880	-0,00496	0,0000245949
2	Siantar Marimbun	0,10189	0,15460	-0,05271	0,0027786719
3	Siantar Selatan	0,10988	0,12700	-0,01712	0,0002932392
4	Siantar Barat	0,17592	0,14690	0,02902	0,0008419833
5	Siantar Utara	0,21565	0,25150	-0,03585	0,0012852670
6	Siantar Timur	0,18690	0,17610	0,01080	0,0001167415
7	Siantar Martoba	0,18758	0,17740	0,01018	0,0001035380
8	Siantar Sitalasari	0,14630	0,08670	0,05960	0,0035525815
				Total SSE	0,0089966174
				MSE	0,0001108768

Table 6. Testing with Models 3-45-1

No	Sub-District	Target	Output	Error	SSE	Results
1	Siantar Marihat	0,11364	0,12850	-0,01486	0,0002209523	1
2	Siantar Marimbun	0,10159	0,16290	-0,06131	0,0037590104	1
3	Siantar Selatan	0,10914	0,13590	-0,02676	0,0007163364	1
4	Siantar Barat	0,17576	0,17360	0,00216	0,0000046710	1
5	Siantar Utara	0,21470	0,27800	-0,06330	0,0040070017	1
6	Siantar Timur	0,18602	0,20210	-0,01608	0,0002585086	1
7	Siantar Martoba	0,18714	0,20750	-0,02036	0,0004145288	1
8	Siantar Sitalasari	0,14609	0,10670	0,03939	0,0015516582	0
				Total SSE	0,0109326675	**88%**
				MSE	0,0012355953	

3.3. Determination of the Best Architectural Model

After training and testing data on models 3-25-1, 3-30-1, 3-45-1, 3-54-1 and 3-68-1 using the help of Matlab and Microsoft Excel tools, the best architectural model is obtained 3-45-1 with an accuracy level of 88% or the highest accuracy compared to the other 4 models. The overall results of the 5 architectural models used can be seen in table 7 below.

Table 7. Comparison of Overall Results of the Architectural Model Used

No	Model	Training			Testing	
		Epoch	Time	MSE	MSE	Accuracy
1	3-25-1	590	00:09	0,0010001080	0,0019661883	33%
2	3-30-1	344	00:06	0,0009978908	0,0019531545	44%
3	**3-45-1**	**553**	**00:09**	**0,0001108768**	**0,0012355953**	**88%**
4	3-54-1	113	00:03	0,0009951149	0,0014581709	44%
5	3-68-1	304	00:05	0,0009935921	0,0036915432	56%

4. Conclusion

a. From these 5 architectural models, after analysis, the 3-45-1 model was chosen as the best model with epoch 553, MSE training 0,0001108768, MSE testing 0.0012355953 and an accuracy rate of 88%.
b. Compared to previous research that uses bipolar sigmoid (tansig) activation function, gradient descent (traingd) and 2 hidden layer training functions, it turns out that in this study using 1 hidden layer, binary sigmoid (logsig) activation function and Gradient descent with momentum training function and adaptive learning rate backpropagation (traingdx) is faster in the network training process, the accuracy obtained is also quite good.

References

[1] M. K. Z. Sormin, P. Sihombing, A. Amalia, A. Wanto, D. Hartama, and D. M. Chan, "Predictions of World Population Life Expectancy Using Cyclical Order Weight / Bias," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.

[2] A. Wanto, A. P. Windarto, D. Hartama, and I. Parlina, "Use of Binary Sigmoid Function And Linear Identity In Artificial Neural Networks For Forecasting Population Density," *International Journal Of Information System & Technology*, vol. 1, no. 1, pp. 43–54, 2017.

[3] S. Setti, A. Wanto, M. Syafiq, A. Andriano, and B. K. Sihotang, "Analysis of Backpropagation Algorithms in Predicting World Internet Users," *Journal of*

Physics: Conference Series, vol. 1255, no. 1, pp. 1–6, 2019.

[4] T. Afriliansyah *et al.*, “Implementation of Bayesian Regulation Algorithm for Estimation of Production Index Level Micro and Small Industry,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.

[5] B. Febriadi, Z. Zamzami, Y. Yunefri, and A. Wanto, “Bipolar function in backpropagation algorithm in predicting Indonesia’s coal exports by major destination countries,” *IOP Conference Series: Materials Science and Engineering*, vol. 420, no. 012087, pp. 1–9, 2018.

[6] M. Fauzan *et al.*, “Epoch Analysis and Accuracy 3 ANN Algorithm Using Consumer Price Index Data in Indonesia,” in *Conference Paper*, 2018, pp. 1–7.

[7] S. P. Siregar and A. Wanto, “Analysis of Artificial Neural Network Accuracy Using Backpropagation Algorithm In Predicting Process (Forecasting),” *International Journal Of Information System & Technology*, vol. 1, no. 1, pp. 34–42, 2017.

[8] A. Wanto *et al.*, “Levenberg-Marquardt Algorithm Combined with Bipolar Sigmoid Function to Measure Open Unemployment Rate in Indonesia,” in *Conference Paper*, 2018, pp. 1–7.

[9] A. Wanto *et al.*, “Analysis of Standard Gradient Descent with GD Momentum And Adaptive LR for SPR Prediction,” 2018, pp. 1–9.

[10] N. Nasution, A. Zamsuri, L. Lisnawita, and A. Wanto, “Polak-Ribiere updates analysis with binary and linear function in determining coffee exports in Indonesia,” *IOP Conference Series: Materials Science and Engineering*, vol. 420, no. 012089, pp. 1–9, 2018.

[11] E. Siregar, H. Mawengkang, E. B. Nababan, and A. Wanto, “Analysis of Backpropagation Method with Sigmoid Bipolar and Linear Function in Prediction of Population Growth,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.

[12] Sumijan, A. P. Windarto, A. Muhammad, and Budiharjo, “Implementation of Neural Networks in Predicting the Understanding Level of Students Subject,” *International Journal of Software Engineering and Its Applications*, vol. 10, no. 10, pp. 189–204, 2016.

[13] 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. 1–17, 2017.

[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] A. Wanto, M. Zarlis, Sawaluddin, and D. Hartama, “Analysis of Artificial Neural Network Backpropagation Using Conjugate Gradient Fletcher Reeves in the Predicting Process,” *Journal of Physics: Conference Series*, vol. 930, no. 1, pp. 1–7, 2017.

[16] G. W. Bhawika *et al.*, “Implementation of ANN for Predicting the Percentage of Illiteracy in Indonesia by Age Group,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.

[17] W. Saputra, P. Poningsih, M. R. Lubis, S. R. Andani, I. S. Damanik, and A. Wanto, “Analysis of Artificial Neural Network in Predicting the Fuel Consumption by Type of Power Plant,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–5, 2019.

[18] Budiharjo, T. Soemartono, A. P. Windarto, and T. Herawan, “Predicting Tuition Fee Payment Problem using Backpropagation Neural Network Model,” *International Journal of Advanced Science and Technology*, vol. 120, pp. 85–96, 2018.

[19] A. Wanto *et al.*, “Analysis of the Accuracy Batch Training Method in Viewing

Indonesian Fisheries Cultivation Company Development," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.

[20] A. Wanto *et al.*, "Analysis of the Backpropagation Algorithm in Viewing Import Value Development Levels Based on Main Country of Origin," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.

[21] I. S. Purba *et al.*, "Accuracy Level of Backpropagation Algorithm to Predict Livestock Population of Simalungun Regency in Indonesia Accuracy Level of Backpropagation Algorithm to Predict Livestock Population of Simalungun Regency in Indonesia," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.

[22] P. Parulian *et al.*, "Analysis of Sequential Order Incremental Methods in Predicting the Number of Victims Affected by Disasters," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.

[23] W. Saputra, J. T. Hardinata, and A. Wanto, "Implementation of Resilient Methods to Predict Open Unemployment in Indonesia According to Higher Education Completed," *JITE (Journal of Informatics and Telecommunication Engineering)*, vol. 3, no. 1, pp. 163–174, 2019.

[24] Budiharjo, T. Soemartono, A. P. Windarto, and T. Herawan, "Predicting School Participation in Indonesia using Back-Propagation Algorithm Model," *International Journal of Control and Automation*, vol. 11, no. 11, pp. 57–68, 2018.

[25] M. R. Lubis, W. Saputra, A. Wanto, S. R. Andani, and P. Poningsih, "Analysis of Artificial Neural Networks Method Backpropagation to Improve the Understanding Student in Algorithm and Programming," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.

[26] S. Setti and A. Wanto, "Analysis of Backpropagation Algorithm in Predicting the Most Number of Internet Users in the World," *JOIN (Jurnal Online Informatika)*, vol. 3, no. 2, pp. 110–115, 2018.

Authors

1st Author
Marseba Situmorang
Student of STIKOM Tunas Bangsa Pematangsiantar
marsebasitumorang465@gmail.com

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

3rd Author
Zulaini Masruro Nasution
Lecturer of STIKOM Tunas Bangsa Pematangsiantar
zulaini@amiktunasbangsa.ac.id