Journal of Telematics and Informatics (JTI)
Vol.6, No.3, September 2018, pp. 238~247
ISSN: 2303-3703

Heuristic Techniques on Weight Optimization of Backpropagation Neural Network

Khoiriya Latifah [a,*], Ika Novita Dewi [b], Ika Ayu Amelia P [c], Richardus Anggi P [d]
[a] *Informatics Department, PGRI Semarang University , Indonesia*
Main Campus : 24 Sidodadi Timur Street Dr. Cipto Semarang
[b] *Informatics Department, Dian Nuswantoro University, Indonesia*
Main Campus: 207 Imam Bonjol Street +6224-3517261
c, Informatics Department Of Dian Nuswantoro University, Indonesia
Main Campus: 207 Imam Bonjol Street +6224-3517261
[d] *Informatics, Dian Nuswantoro University, Indonesia*
khoiriyalatifah@upgris.ac.id, ikadewi@research.dinus.ac.id, amelianid@yahoo.com, ricardus.anggi@dsn.dinus.ac.id

Abstract

The number of visitors the immigration office who made passport or passport renewal every month is very volatile, this problems make the Immigration Office maintain quality of service, thus requiring prediction methods when there will be a surge of visitors so that the quality of service is maintained. The immigration office should have some information to make predictions. Perfect information will make it easier , good predictions and accurate predictions. The exact method is one of the method predict accurately. Its called artificial neural network that a computational method mimics system of neural network biology. Artificial neural networks are formed to solve a particular problem such as pattern recognition or classification because of the learning process.This study uses Heuristic Backpropagation to increase the speed of the training process of neurons in making predictions.

Keywords*: Neural Network, Backpropagation, Heuristik , Prediction*

1. Introduction

The fluctuating number of visitors immigration office can make problems. To maintain public quality of service, Semarang Immigration Office requires a prediction method to predict when there will be a surge of visitors than the quality of service is maintained. Prediction as one of the good techniques to find out how many visitors of immigration office that make quality of service is maintained. [1]. Decision in predicting is very important, because every mistake in the prediction will make the image of the immigration office decrease. For this reason the immigration office must have some information to make predictions. Perfect information will facilitate predictions and good predictions are accurate predictions. To make predictions accurately, the right method is also needed.

Prediction algorithm is a forecasting technique that shows fewer error techniques compared to statistical methods [2] [3]. Neural Networks are artificial neural networks that can think like humans, have the ability to learn and the learning process is supervised by inputs resulting in network output capable of solving a problem such as pattern recognition, classification or prediction. [4] [5]

Neural Network is a classification algorithm used to assess data objects and enter data into certain classes. Classification is training or learning on the function f (target / label) that maps each attribute x to one of the number of available class y labels. In the classification, a model is developed where the model is used to classify or predict an object to be known in which class the data object is in the saved model [6].

One simple Neural Network algorithm is Backpropagation. Backpropagation is a supervised learning algorithm and is usually used by multiple layers of perceptrons to change the weights connected to the neurons in the hidden layer. The basic principle of a simple backpropagation algorithm is to fix the network weights by making the direction of the

performance function decrease rapidly [7]. One method that is commonly used to implement repairs to weight values is the Heuristic method. The Heuristic Method includes the method used to speed up the training process and the improvement of the weighting value is carried out every data input operation.

From the background above, this study uses the Backpropagation Neural Network algorithm with Heuristic method to improve better performance as an approach in predicting the number of visitors to the Immigration office. To correct the weights based on gradient descent with adaptive learning rates that are adaptive compared to the resilent method to improve input weights.

2. Research Method

To make predictive performance improvements the number of visitors to the Immigration office. in Backpropagation Neural Network the writer uses Heuristic method.

Research methods to solve problems as follows:

1. Analysis of Problems
   Find problems in the field. In this study the problem faced was the fluctuating number of passport applicants
2. Data collection
   Data comes from the Immigration office database in 2014 – 2015
3. Data Experiment Preparation
   Preprocessing data is normalizing data derived from secondary data for passport applicants.
4. Experiment
   Data experiments used Matlab and Rapid Miner tools to conduct experiments.
5. Training and Testing
   Secondary data after being normalized are divided into two, namely training data and test data.
6. If training and testing fail then data modification and data format are modified. If training and testing are successful, it will be forwarded to validation and verification
7. Validation and Verification
   Validation and Verification using MRSE to determine the error level of the test results.
8. Discussion
   After validation and verification, further discussion and writing of the report is carried out.

Problem Analysis

To maintain the credibility and quality of service, the immigration office needs to maintain service quality because people who are increasingly critical and the people's lifestyles are now practical so that it takes calculations in forecasting the number of passport makers to avoid long lines and the immigration office can arrange employee leave schedules or add infrastructure in the event of a passport maker surge

3. Results and Analysis

3.1. Data Collection

The data used in the study were data on the arrival of visitors to the Immigration Office in Semarang in 2013. For the data used was data every month from 2013 to 2014 concerning the number of applicants passing the Semarang Immigration Office. The data used is 24 months, the data used for transmission is data from 2013, the first 17 months, while testing uses 2014 data for 4 months.

Tabel 1. Data Univariate and Multavariate

a. Training data Univariate

Mei 2013	8730
Juni 2013	9786
Juli 2013	8438
Agust 2013	10105
Sept 2013	4618
Okt 2013	4615
Nov 2013	5220
Des 2013	5005
Jan 2014	3962
Feb 2014	4342
Maret 2014	4051
Apr 2014	4009
Mei 2014	5797
Juni 2014	6526
Juli 2014	6684
Agust 2014	4814
Sept 2014	3520
Okt 2014	3849
Nov 2014	5048
Des 2014	4223

b. Data Multivariate

Date	Jumlah
Jan 2014	5574
Feb 2014	4373
Maret 2014	4762
Apr 2014	5701
Mei 2014	8730
Juni 2014	9786
Juli 2014	8438

Date	Y+1	Y+2	Y+3	Y+4
Jan 2014	5574	4373	4762	5701
Feb 2014	4373	4762	5701	8730
Maret 2014	4762	5701	8730	9786
Apr 2014	5701	8730	9786	8438

3.2. Heuristic Method OF BPNN

To obtain the best RMSE results, the Heuristic method requires setting parameters for the BPNN method. Among other things, Learning Parameters, namely maximum epoch, error, training and learning rate functions. These parameters will affect the system performance in the learning process of the prediction process. For maximum epoch, do epoch testing to get the best epoch value. The training function used consists of 2 types, namely the Matlab default function, namely Gradient Descent with Adaptive Learning Rate (traingda), and Resilent Backpropagation (trainrp). While the learning rate is given an initial value of 0.1 to 1.

Determining the Function Weight that will be used based on the activation function of determining the minimum and maximum weight values is determined on the basis of the value of the function range, namely the binary sigmoid function and the bipolar sigmoid function. Because in this study using Matlab software, the weight will be determined automatically randomly by Matlab

3.3. Pre Processing and Processing Data

In order for variables to be processed by ANN, it is necessary to normalize these variables so that the variables are in the range [0 - 1]. Besides that, you need to change the vector of the variable to variabell line. This normalization is needed to avoid missing values, scale the attribute values of a data so that it can be a value in a certain range, besides that it aims to balance the attribute values of different dimensions, normalize data with a range between 0 to 1, so that the equation:

$$N = \frac{0{,}8\ (D - D_{min})}{D_{max} - D_{min}} + 0{,}1$$

N = Normalized Data
D = data that will be normalized
D_{min} = the smallest data in a range
D_{max} = the largest data in a range

Tabel 2. Normalized Data

x-4	x-3	x-2	x-1	x-t
0,24013	0,74691	0,1	0,9	0,16564
0,85349	0,65695	0,9	0,1	0,9
0,65709	0,9	0,10044	0,1	0,1
0,9	0,10044	0,1	0,18816	0,1
0,10397	0,1	0,69603	0,6157	0,18816
0,51526	0,9	0,76328	0,1	0,6157
0,9	0,76328	0,1	0,34165	0,1
0,9	0,1	0,39147	0,16826	0,34165
0,1	0,9	0,28757	0,198995	0,16826
0,24899	0,11879	0,1	0,9	0,198995
0,11335	0,1	0,6683	0,9	0,9
0,1	0,63473	0,85275	0,9	0,9
0,16839	0,21912	0,23011	0,1	0,9
0,86005	0,9	0,42718	0,1	0,1
0,9	0,42718	0,1	0,18319	0,1
0,22082	0,1	0,13072	0,24267	0,18319
0,1	0,1	0,9	0,16564	0,00000

x-1	x-2	x-3	x-4	x-t
0,823	0,1	0,334	0,9	0,823
0,1	0,171	0,343	0,9	0,1
0,1	0,249	0,674	0,9	0,1
0,1	0,693	0,9	0,636	0,1

Based on the design of the research steps above, each data will be processed using the five methods above.

After data normalization, it is then converted into training data which functions to determine the amount of input. The following is the experimental results of the model of the neural network method measured using Mean Square Error. Determination of parameters in the neural network is used to find the best model for predicting, namely by finding the best value of each parameter. Determination of parameters in the neural network in this study is based on, Epoch or training cycle, Learning Rate, the number of neurons in the Hidden Layer and the train method used.

a. Parameter Analysis

Parameter analysis will be carried out for the Neural Network parameter as a new method used. Parameter analysis will provide reasons and descriptions of why these parameters are used as well as the results of forecasting generated by the use of these parameters.The parameters that influence the use of the Neural Network are functions. Epoch or training cycle, Learning Rate, the number of neurons in the Hidden Layer and the train method used.

The input layer will be varied with input from 2 to 4 input layers. Distribution of input data for training is as follows:

- ✓ Input layer = 2, then the data set input is January to December 2013 with the 2015 data target (xt-2).
- ✓ Input layer = 3, then the data set input is used from January to December 2014 with the 2015 data target (xt-3).
- ✓ Input layer = 4, then the data set input is used from January to December 2014 with the 2015 data target (xt-4).

The parameters that must be set are :

- Maximum Epoh : net.trainParam.epochs = MaxEpoh

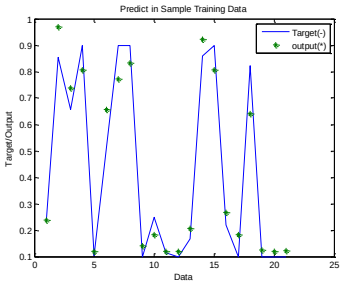
- Performance goals : set.trainParam.goal = TargetError
- LearningRate ; net.trainParam.lr = LearningRate
- Ratio of increasing Learning Rate : net.trainParam.lr_inc = IncLearningRate
- Ratio of decreasing Learning Rate : net.trainParam.lr_decc = DecLearningRate
- Maximum failure : net.trainParam.max_fail = MaxFail
- Maximum of increasing performance : net.trainParam.max_perf_inc = MaxPerfInc
- Gradient Minimum : net.trainParam.min_grad = MinGradient

b. Forecasting Result Analysis

To obtain the value of training cycles in this study determined by testing the training data by 80% where the remaining 20% as data testing. In training, the best model is needed by entering values manually in the range of 100 to 1000 for training cycles, the learning rate is 0.1 with the number of neurons in the hidden layer as much as 3. The following are the results of experiments that have been conducted to determine the value of training cycles.

Tabel 3. Amount Determination of *Training Cycle (Epoch)* based on *Training Method Gradient Descent Backpropagation with Adaptive Learning Rate (traingda)* 4 layer input .

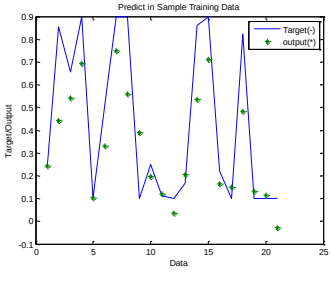
Epoch	Learning Rate	Size	Loop 1	Loop 2	Loop 3	Average
			MSE			
100	0.1	3	0.0198	0.0584	0.0095	0.0280
200	0.1	3	0.0833	0.0099	0.0292	0.0408
300	0.1	3	0.0075	0.0151	0.0076	0.0100
400	0.1	3	0.0291	0.0113	0.0386	0.0263
500	0.1	3	0.0139	0.0232	0.0075	0.0148
600	0.1	3	0.0444	0.0171	0.0066	0.0227
700	0.1	3	0.0321	0.0216	0.0199	0.0245
800	0.1	3	0.0167	0.0092	0.0180	0.0146
900	0.1	3	0.0122	0.0120	0.0228	0.0156
1000	0.1	3	0.1000	0.0148	0.0097	0.0415



In the test results in the table above shows that the training cycle that produces the smallest MSE value is at the 300th epoch.

Tabel 4. Amount Determination of Training *Cycle (epoch)* base on Training *Method Resilient Backpropagation* with 4 input layer.

Epoch	Learning Rate	Size	Loop 1	Loop 2	Loop 3	Average
			MSE			
100	0.1	3	0.0238765	0.00978711	0.0479172	0.02719
200	0.1	3	0.0115274	0.0110453	0.0241125	0.01556
300	0.1	3	0.0160088	0.090164	0.0119615	0.03938
400	0.1	3	0.00992784	0.0202187	0.0170025	0.01572
500	0.1	3	0.0182468	0.0131517	0.0105296	0.01398
600	0.1	3	0.0536098	0.0168425	0.0134157	0.02796
700	0.1	3	0.035263	0.015259	0.0148696	0.0218
800	0.1	3	0.018262	0.0154477	0.0303097	0.02134
900	0.1	3	0.0501091	0.0108301	0.0493871	0.03678
1000	0.1	3	0.0180692	0.00962202	0.00793606	0.01188



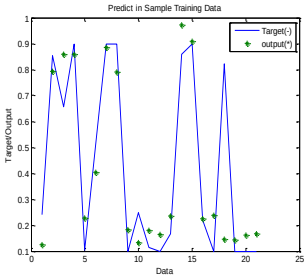
The training cycle value is selected based on the smallest MSE value generated. In the test results in the table above shows that the training cycle that produces the smallest MSE value is 1000. So the use of training cycles in this study refers to the results of the epoch produced by the Training Method Gradient Descent Backpropagation with Adaptive Learning Rate (traingda) and Resilent Method BPNN.

Learning Rate

Determine the value of learning rate by conducting a test to enter a value in the range of 0.1 to 1, and the training cycle value from the previous experiment. The following are the results of experiments that have been carried out for the determination of learning rate:

Tabel 5. Value Determination *Learning Rate* base on Training *Method Gradient Descent Backpropagation* with 4 input layer.

Epoch	Learning Rate	Size	Loop 1	Loop 2	Loop 3	Average
			MSE			
300	0.1	3	0.0187833	0.0740677	0.0206939	0.03785
300	0.2	3	0.0145497	0.0129471	0.0666041	0.03137
300	0.3	3	0.0297808	0.0146479	0.0131909	0.01921
300	0.4	3	0.0106444	0.161492	0.0260278	0.06605
300	0.5	3	0.094244	0.0386587	0.0468801	0.05993
300	0.6	3	0.0214462	0.148932	0.0726079	0.0810
300	0.7	3	0.0114496	0.0223068	0.0871996	0.04032
300	0.8	3	0.01514	0.106767	0.0118445	0.04458
300	0.9	3	0.121083	0.0416242	0.0130355	0.05858
300	1	3	0.0166704	0.0489444	0.0621551	0.04259



In the test results in the table above shows that the learning rate that produces the smallest MSE value is 0.3 with a training cycle value of 300.

Table 6. Value Determination of *Learning Rate* base on *Training Method Resilient Backpropagation* with 4 input layer

Epoch	Learning Rate	Size	Loop 1	Loop 2	Loop 3	Average
			MSE			
1000	0.1	3	0.0162278	0.0149548	0.0217547	0.017646
1000	0.2	3	0.0910096	0.00969648	0.0217414	0.040816
1000	0.3	3	0.0139632	0.00721348	0.0233886	0.014855
1000	0.4	3	0.0279929	0.0309764	0.0435498	0.034173
1000	0.5	3	0.0496665	0.0117574	0.142404	0.067943
1000	0.6	3	0.0113476	0.00882756	0.0705692	0.030248
1000	0.7	3	0.0151781	0.0206298	0.0238986	0.019902
1000	0.8	3	0.014911	0.00968918	0.00819959	0.010933
1000	0.9	3	0.0154996	0.0257577	0.0181548	0.019804
1000	1	3	0.00915606	0.0504132	0.0114855	0.023685

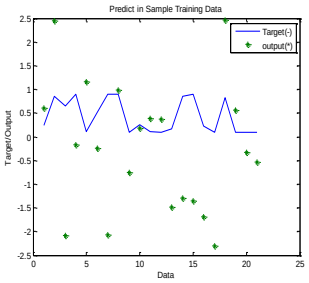
The learning rate value is selected based on the smallest MSE value produced. In the test results in the table above shows that the learning rate that produces the smallest MSE value is 0.8 with a 1000 cycle training value.

Determination of the Number of Neorun in the Hidden Layer

In this experiment the determination of the number of hidden layers is 1 and to determine the number of neurons in the hidden layer by conducting a manual test to enter values in the range 1 to 10 and multiples of 10, using the number of training cycles, the learning rate value from the previous experiment. The following are the results of experiments that have been conducted to determine the number of neurons in the hidden layer.

Tabel 7. Determination of the Number of Neurons in the Hidden Layer based on the Gradient Descent Backpropagation Training Method with 4 input layers.

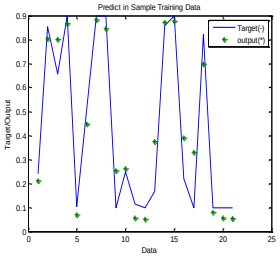
Epoch	Learning Rate	Size	Loop 1	Loop 2	Loop 3	Average
			MSE			
300	0.3	1	0.0448263	0.00539225	0.00729711	0.019172
300	0.3	2	0.0867457	0.0158321	0.0479035	0.05016
300	0.3	3	0.0154393	0.00897234	0.0281491	0.01752
300	0.3	4	0.022414	0.0175126	0.0222558	0.020727
300	0.3	5	0.0158033	0.0133544	0.0120125	0.013723
300	0.3	6	0.0657681	0.0211031	0.00989279	0.032255
300	0.3	7	0.0079746	0.0914202	0.0324167	0.043937
300	0.3	8	0.0204719	0.0308988	0.0153719	0.022248
300	0.3	9	0.169865	0.0243473	0.0597412	0.084651
300	0.3	10	0.0320226	0.0341095	0.0259409	0.030691
300	0.3	20	0.0318108	0.027274	0.146512	0.068532
300	0.3	30	0.0650632	0.07829	0.154291	0.099215
300	0.3	40	0.0333403	0.088802	0.10509	0.075744
300	0.3	50	0.0450168	0.312693	0.105534	0.154415
300	0.3	60	0.141707	0.0886637	0.228333	0.152901
300	0.3	70	0.266693	0.220816	0.271	0.252836
300	0.3	80	0.117369	0.36401	0.269036	0.250138
300	0.3	90	0.423918	0.133972	0.691975	0.416622
300	0.3	100	0.597542	0.81876	1.30918	0.908494



The number of neurons in the best hidden layer is selected based on the smallest MSE value generated. In the test results in the table above shows that the number of neurons in the hidden layer that produces the smallest MSE value is 5.

Tabel 8 . Determination of the Number of Neurons in the Hidden Layer based on the Training Method Resilient Backpropagation with 4 input layers.

Epoch	Learning Rate	Size	Loop 1	Loop 2	Loop 3	Average
			MSE			
1000	0.8	1	0.0720471	0.122116	0.00698424	0.067049
1000	0.8	2	0.0148577	0.00373831	0.0189475	0.012515
1000	0.8	3	0.0190198	0.0149408	0.0130469	0.015669
1000	0.8	4	0.0569433	0.0172525	0.0289811	0.034392
1000	0.8	5	0.0218258	0.00838578	0.0491046	0.026439
1000	0.8	6	0.00895714	0.0273353	0.0130344	0.016442
1000	0.8	7	0.0145935	0.0507736	0.0189263	0.028098
1000	0.8	8	0.051546	0.0181881	0.0165795	0.028771
1000	0.8	9	0.0423542	0.0309807	0.0174921	0.030276
1000	0.8	10	0.0180155	0.0485161	0.0113857	0.025972
1000	0.8	20	0.0965	0.106862	**0.0739231**	0.092428
1000	0.8	30	0.254178	0.0419448	0.084854	0.126992



1000	0.8	40	0.225266	0.0525401	0.239463	0.172423
1000	0.8	50	0.0958612	0.0853429	0.181136	0.12078
1000	0.8	60	0.198968	0.33809	0.293478	0.276845
1000	0.8	70	0.441107	0.183834	1.03599	0.553644
1000	0.8	80	0.282911	0.735855	0.501406	0.506724
1000	0.8	90	1.12094	0.219868	0.552479	0.631096
1000	0.8	100	0.609997	0.778693	0.624912	0.671201

The number of neurons in the best hidden layer is selected based on the smallest MSE value generated. In the test results in the table above shows that the number of neurons in the hidden layer that produces the smallest MSE value is 2.

From the results of the initial testing using training data up to 4 periods (xt-4) or backward 4 years before, as a reference comparison of the target position (xt), using the Gradient Descent algorithm obtained the best parameter value at epoch position = 300, learning rate = 0.3 and the number of neurons in the hidden layer = 5. With the RMSE value = 0.018843 and the R value approaching 1 that is as much as 0.9.

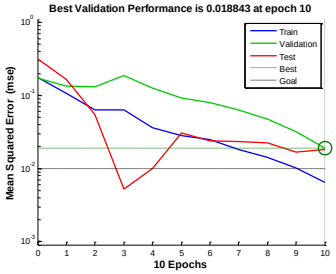


Figure 1. RMSE value on Gradient Descent Adaptive Learning Rate

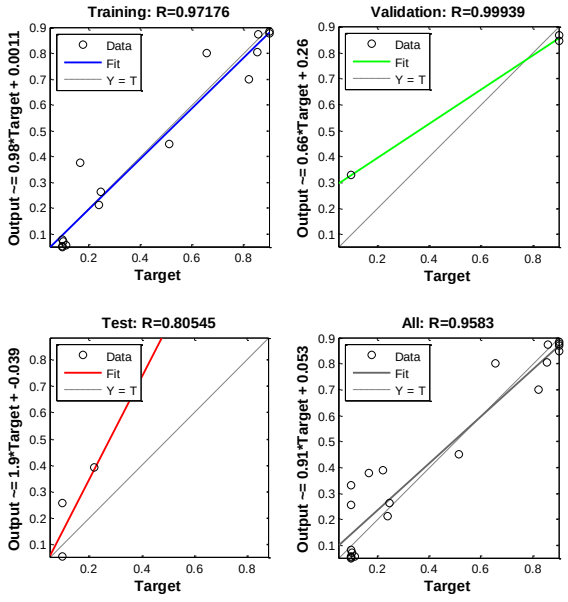


Figure 2. R value on Gradient Descent Adaptive Learning Rate

While the Resilient Backprogration algorithm obtained the best parameter values at epoch = 1000, learning rate = 0.8 and the number of neurons in the hidden layer = 2. The value of RMSE = 0.006 and R = 0.9

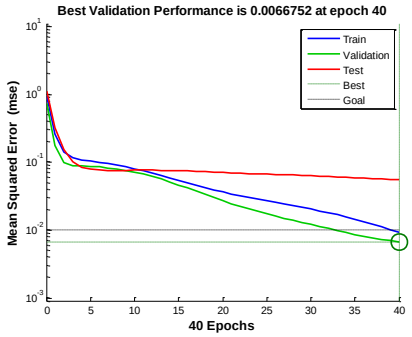


Figure 3. RMSE value on Resilent Backpropagation

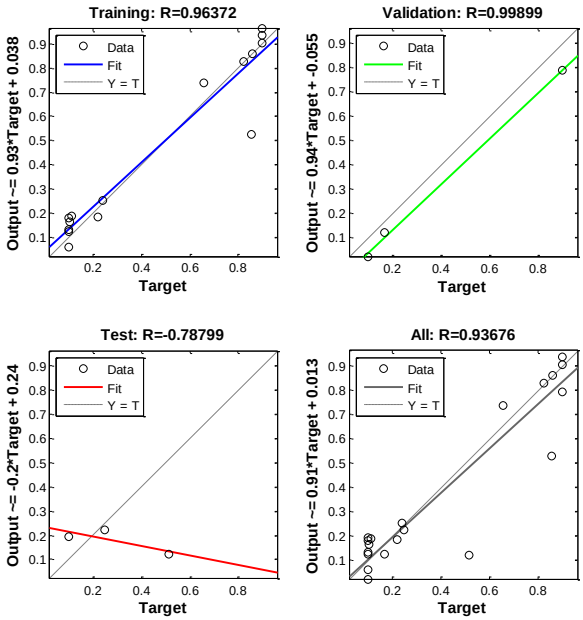


Figure 4.10. R value on Resilent Backpropagation

4. Conclusion

Based on the experimental results, starting from the initial stage until evaluation, it can be concluded that the rentet prediction model when making passports using Bacpropagatian-based Neural Network method is quite accurate with the 2-9-1 network architecture, namely 1 input layer with 2 neurons, 1 hidden layer with 9 neurons and 1 out
put layer with 1 neuron. While the use of gradient descent with adaptive learning rate (traingda) obtains the best parameter value at epoch = 300, learning rate = 0.3 and the number of neurons in hidden layer = 5. With the RMSE value = 0.018843 and R value approaching 1 that is 0.9.
As for the use of Resilient Backprogration (trainrp) gradient, the best parameter values at epoch = 1000 are obtained, learning rate = 0.8 and the number of neurons in the hidden layer = 2. RMSE = 0.006 and R = 0.9
Thus the use of resilent backpropagation gradient descent has a smaller RMSE compared to simple backpropagation and adaptive learning rate backpropagation.

The research will be more accurate if there are more training sets and for future research the prediction system to determine which visitors to the Immigration office can combine neural Network methods with other methods so that the results will be more accurate.

References

[1]. Bogdan Oancea, Stefan C. , *Time Series Forecasting Using Neural Network*, 2015, Challenges of The Knowledges Society IT in Social Sciences

[2]. Cao Hongmei, Han Feng, 2014, *The Urban Arterial Traffic Flow Forecasting Based On BP Neural Network*, Fourth International Conference on Instrumentation and Measurement, Computer, Communication and Control, China

[3]. Liang Lanzhen, ShaoFan, 2010, *The Study On Short Time Wind Speed Prediction Based On Time Series Neural Network Algorithm*, IEEE.

[4]. Mirza Cilimkovic, , 2015, Neural Network and Backpropagation Algorithm, Institute of Technology Blanchardstown, Blanchardstown Road North Dublin 15, Ireland.

[5]. J. Kamruzzaman and R.A. Sarker. 2009. "Comparing ANN Based Models with ARIMA for Prediction of Forex Rates," vol. 22, pp. 2-11.

[6] Eko P, 2012, Data Mining Konsep dan Aplikasinya Menggunakan Matlab, Yogyakarta, Andi Publisher.

[7]. Sri Kusumadewi, 2004, Membangun Jaringan Syaraf Tiruan Dengan Matlab, Yogyakarta, Graha Ilmu.

[8] Tjiptono, Fandy. 2005. Prinsip Prinsip Total Quality Service. Yogyakarta : Penerbit Andi

[9]. Haykin, S., 1994, *Neural Networks: A Compr, Jurnal Universitas Panca Marga ehensive Foundation*, MacMillanCollege Pub lishing Company.

[10]. Kusumadewi, S., 2003, *Artificial Intelligence* (*Teknik dan Aplikasinya*), Graha Ilmu, Yogyakarta.

[11]. Kusumadewi, S., 2004, *Membangun Jaringan Syaraf Tiruan*, Graha Ilmu, Yogyakarta.

[12]. Onwubolu, G. C., 1999, *Manufacturing Features Recognition using Backpropagation Neural Networks*, Journal of Intelligent Manufacturing Volume 10.

[13]. Ozel, T., Nadgir, A., 2001, *Prediction of flank wear by using back propagation neural network modeling when cutting hardened H-13 steel with chamfered and honed CBN tools*, International Journal of Machine Tools & Manufacture, Cleveland.

[14] QianZhang and Tongna Liu, "A Fuzzy Rules and BACKPROPAGATION Neural Network Method for Mid-long-term Electric Load Forecasting," in Second International Conference on Computer and Network Technology, 2010, pp. 442-446.

[15] Gao Guorong and Liu Yanping, "Traffic Flow Forecasting based on PCA and BACKPROPAGATION Neural Network," in 2010 International Conference of Information Science and Management Engineering, 2010, pp. 158-161.