



# Forecasting Time Series Data Using Bayesian Regularization

Doni El Rezen Purba<sup>1</sup>, Dewi Sartika Br Ginting<sup>2</sup>

<sup>1</sup>Informatics Engineering Department, Universitas Katolik Santo Thomas, Medan, Indonesia,

<sup>2</sup>Computer Science Departments, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia.

## Article Info

### Article history:

Received Apr 27, 2021

Revised Mei 27, 2021

Accepted Jun 28, 2021

### Keywords:

Backpropagation,  
Bayesian Regularization,  
Forecasting,  
Neural Network.

## ABSTRACT

Forecasting or predicting future events is important to take into account in order for an activity to proceed properly. Flights predict the weather forecast, the banking industry predicts the price of currency, the health world predicts the disease, the retail business predicts total sales. prediction or forecasting of events is calculated using past data, usually in the form of time series. Artificial neural networks are capable of forecasting time-series data. Forecasting results with artificial neural network is influenced from the network architecture model is determined, one of which determination of training function. This study uses the bayesian regularization training function to forecast time clock data with several layer count models and the number of neurons. The results obtained with the number of 3 layers and each neuron of 36, 12, 6 for the best process performance, and the number of neurons 24, 12, 6 for the shortest iteration process.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## Corresponding Author:

Doni El Rezen Purba,  
Informatics Engineering Department, Universitas Katolik Santo Thomas, Medan, Indonesia,  
Jalan Setia Budi No.479 F, Tanjung Sari Medan.  
Email: [donielerzenpurba@gmail.com](mailto:donielerzenpurba@gmail.com)

## 1. INTRODUCTION

Forecasting is a process of quantitative estimation of what will happen in the future based on relevant data in the past [1]. In the field of aviation, for example, weather forecasting to anticipate failure in flight. Predictions of possible severe weather resulting in communication disturbances due to storms, or the presence of cumulonimbus clouds that could endanger the flight. In the retail business is also so, conducted forecasting to estimate the increase or decrease sales of a product in order to be done anticipation to avoid losses.

Many methods can be used to do the forecasting. Can use statistical models or with artificial neural networks [2]. Artificial neural networks are a method of forecasting that is directed at a simple mathematical model of the workings of the human brain. The complex nonlinear relationship between response and predictor variables [3]. The statistical model can be called by classical forecasting methods and artificial neural networks called modern forecasting methods. There is also a combination of both to do the forecasting.

Many research has been done to determine the best model of artificial neural network architecture[4]. Problems about the slow process of training, the variety of existing data, the increasing need for information from available data and the increasingly sophisticated computing equipment to improve the process of computer work [4].

The results of forecasting with neural network motion are influenced based on the data form and the network architecture parameters used. The choice of neural network model must be in accordance with the form of data to be used . [5] conducted a research method of artificial backpropagation network to forecast sales of oral care products ranging from suppliers to final

consumers to help in determining business decisions. [6] used the Convolutional Neural Network (CNN) model on his research on sales forecasts in the E-commerce field using promotional history data, price changes and user preferences to help manage the workforce, cash flow and sources power on the company based on the results of forecasting done.

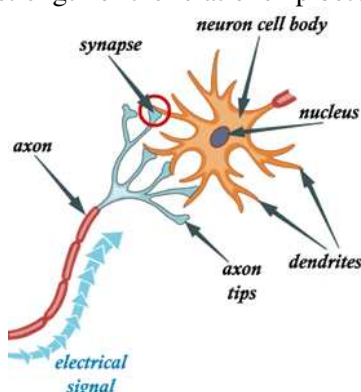
The selection of network models influences the outcomes of the learning process, both the accuracy of the results and the length of the calculation process. Selection of activation functions, training functions and the number of layers and neurons for the training and testing process should be well considered. [4] conducted a study titled Bayesian Regularization in a Neural Network Model to Estimate Lines of Code Using Function Points, stating that, the neural network model trained using Bayesian Regularization gave the best results and was suitable for the study. Then, doing research on the Effects of Training Functions of Artificial Neural Networks (ANN) on Time Series Forecasting, obtained from all the training algorithms used for hourly weather history data forecasting, levenberg marquardt proved to have the least squares error and correlation coefficient [7]. Based on the exposure, in this research will be conducted research for testing the best neural network model against some time series data that has been provided. To get the best forecasting results.

## 2. RESEARCH METHOD

### 2.1. Neural Network

An artificial neural network processes large amounts of information in parallel and distributed, this is inspired by the biological brain work model. Artificial neural systems is a distributed and parallel processed information processing structure, consisting of a processing element (which has local memory and operates with local information) interconnected along with a direct-line flow called a connection [8]. Each processor element has a single outlet connection that fan out to the desired number of collateral connections (each connection carrying the same signal from the output of the processing element). The output of the processing element can be any kind of mathematical equation desired. The entire process that takes place on each processor element must really be done locally, ie the output depends only on the input value at that moment obtained through the connection and the value stored in the local memory.

The structure in Figure. 1 is the basic standard form of a unit of simplified human brain network units. The human brain tissue is composed of  $10^{13}$  neurons connected by about  $10^{15}$  dendrites. The dendrite function is as a transmitter of signals from the neuron to the neurons connected to it. Nucleus is the nucleus of a neuron, the axon acts as the output channel of the neuron, and the synapses that govern the strength of the relationship between neurons.



**Figure 1.** Structure of neural network biology

An artificial neural network consists of a collection of neuron groups arranged in layers :

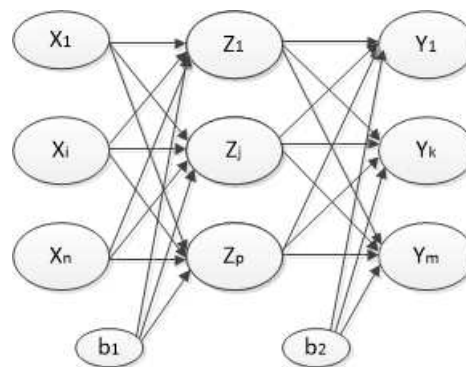
- a. Input Layer : serves as a network link to the outside world (data source).
- b. Hidden Layer : A network can have more than one hidden layer or even can not have it at all.
- c. Output Layer : The working principle of neurons in this layer is similar to the working principle of neurons in the hidden layer and here also the Sigmoid function is used, but the output of the neurons in this layer is considered to be the result of the process.

In general, there are three types of neural networks that are often used based on the type of network it is :

- Single-Layer Neural is a neural network that has connections on its input directly to the output network.
- Multilayer Perceptron Neural Network is a neural network that has a layer called "hidden", in the middle of the input and output layers. Hidden is variable, can be used more than one hidden layer.
- Recurrent Neural Networks Neural network is a neural network that has characteristics, namely the existence of a feedback connection from the output to the input.

## 2.2. Backpropagation Neural Network

Backpropagation is one of the training methods of Artificial Neural Network. Backpropagation uses a multilayer architecture with supervised training methods. The Back Propagation Model has several units that exist in one or more hidden layers. Figure 3 is a back propagation model architecture with  $n$  input (plus one bias), a hidden layer consisting of  $p$  units (plus a bias) and  $m$  units of output units.



**Figure 2.** Architecture of Backpropagation Neural Network

At the feed-forward stage, each input unit ( $X_i$ ) receives the input signal and sends this signal to each hidden unit  $Z_1, \dots, Z_p$ . Each hidden unit counts its activation and sends its signal ( $Z_j$ ) to each output unit. Each output unit ( $Y_k$ ) calculates its activation ( $Y_k$ ) to indicate the network response to the given input pattern.

During the training, for each unit of output compared to  $Y_k$  activation with the target  $T_k$  to determine the error between the input pattern and the output unit. After obtained error value, factor  $\delta_k$  ( $k=1, \dots, m$ ) calculated  $\delta_k$  which is used to distribute the error on the  $Y_k$  output unit back to all units on the previous layer (hidden unit connected to  $Y_k$ ). Then this error is used to change the weights between the output and the layer with the hidden layer. In the same way, the factor  $\delta_j$  ( $j=1, \dots, p$ ) is calculated for each unit  $Z_j$ . Factor  $\delta_j$  used to change the weights between the hidden layers and the input layer.

After all the factors  $\delta$  are determined, the weights for the whole layer are adjusted directly. The weight adjustment  $W_{jk}$  (from hidden unit  $Z_j$  to  $Y_k$  output unit) is based on the factor  $\delta_k$  and activation of unit  $Z_j$ . The weight adjustment of the  $v_{ij}$  (from the  $X_i$  input unit hidden unit  $Z_j$ ) is based on the factor  $\delta_j$  and activation of the input unit  $x_i$ .

The usual activation function used to train artificial neural network is sigmoid function, both binary and bipolar. Here's the training algorithm [9] :

**Step 0.** Initialize initial weights (specifies a small random value)

**Step 1.** As long as the stop condition is false, take steps 2-9

**Step 2.** For each pair of training, step 3-8 (Feedforward)

**Step 3.** Each input unit ( $X_i$ ,  $i = 1, \dots, n$ ) receives the input signal  $X_i$  and sends the signal throughout the unit on the next layer (hidden layer).

**Step 4.** For each hidden unit ( $Z_j$ ,  $j = 1, \dots, p$ ), the input signal is weighted in and applied the activation function to calculate its output and send this signal to all subsequent layer units (output layer).

**Step 5.** For each Output unit ( $Y_k$ ,  $k = 1, \dots, m$ ) see the weighted input in sequence.

**Back Propagation from error**

**Step 6.** For each output unit ( $Y_k$ ,  $k = 1, \dots, m$ ) received a target pattern corresponding to the input pattern, calculated the error and calculated the weight correction and calculated the correction and sent  $\delta_k$  to the unit in the previous layer.

**Step 7.** For each hidden unit ( $Z_j$ ,  $j = 1, \dots, p$ ), the summed delta function is then multiplied by its activation function to calculate, calculated its weight correction and correction.

**Step 8.** For each output unit ( $Y_k$ ,  $k = 1, \dots, m$ ) the bias and weights are changed ( $j = 1, \dots, p$ ). For each hidden unit ( $Z_j$ ,  $j = 1, \dots, p$ ) the weight and bias are changed ( $i = 1, \dots, p$ ).

### 2.3. Forecasting

Forecasting is a process of quantitative estimation of what will happen in the future based on relevant data in the past [1]. Demand forecast is the level of demand for products that are expected to be realized for a certain period of time in the future.

Basically the approach of forecasting can be classified into two approaches, namely qualitative forecasting and quantitative forecasting [2]. There are 4 types of data patterns in forecasting :

- Trend, the trend data pattern shows the movement of data tends to increase or decrease for a long time.
- Seasonality, seasonal data patterns are formed due to seasonal factors, such as weather and holidays.
- Cycles, occur when variations of corrugated data over a duration of more than one year are influenced by political factors, economic changes (expansion or contraction), known as business cycles.
- Horizontal/Stasionary/Random variation, occurs if the data fluctuates around a random average value without forming a clear pattern such as a seasonal pattern, trend or cycle.

### 2.4. Bayesian Regularization

Regularization plays a role in improving the generalization process by limiting the size of the weight of a network. If the value of network weight is smaller then the network will respond more subtly. With regularization, a large, simplified network must be able to represent the actual function. In the classic Backpropagation algorithm it aims to minimize functionality  $F=E_d$ , where :

$$E_w = \frac{1}{n} \sum_{i=0}^n (w_i)^2 \quad (1)$$

In this case  $n$  is the number of inputs in the training set,  $t_i$  is the target value in the data to- $i$  and  $a_i$  is the output for the data to- $i$  which is obtained as a neural network response. The regulatory method changes the performance of the error function by adding a standard deviation of the weights and biases :

$$F = \beta E_d + \alpha E_w \quad (2)$$

$\alpha\beta$  is regularization parameter, and  $E_w$  define as :

$$E_d = \sum_{i=0}^n (t_i - a_i)^2 \quad (3)$$

$W_i$  is a weight or a threshold. Using the equation to change the error performance function allows the network to obtain the smallest weights and thresholds, but it can not determine effective network weight and thresholds. The conventional method is often difficult to determine the size of the parameter proposes a network that can adjust the size of adaptive parameters using Bayesian theoretical framework, and enables achievement of optimal performance.

## 3. RESULTS AND DISCUSSION

In this study used time series data of the population of a region in Indonesia for the forecasting process. The training function used is Bayesian Regularization. The best network model is determined based on the accuracy and speed of the forecasting process. Prediction will be done by trying some models of the number of layers and the number of neurons. Table 1 below is a form of network model that will be used for forecasting :

**Table 1.** Number of Layers and Neurons

Model	Number of neurons		
	Layer 1	Layer 2	Layer 3
1	12	6	2
2	24	12	6
3	36	12	6
4	48	12	6
5	48	24	12

The program used for forecasting process in this research is matlab. Source code as follows:

```
rng('default')
```

```
inputs = data_latih;
```

```
targets = target_latih;
```

```
x1 = number of neurol layer 1
```

```
x2 = number of neurol layer 2
```

```
x3 = number of neurol layer 3
```

```
net = newff(inputs,targets, { x1, x2, x3 });
```

```
net = train(net,inputs,targets);
```

```
net.trainFcn = 'trainbr';
```

```
outputs = net(inputs);
```

```
errors = outputs - targets;
```

```
perf = perform(net,outputs,targets)
```

```
figure,
```

```
plot(outputs,'bo-')
```

```
hold on
```

```
plot(targets,'ro-')
```

```
hold off
```

```
grid on
```

```
title(strcat(['Plot Performa NNBR, Value = ', num2str(perf)]))
```

```
xlabel('Month -')
```

```
ylabel('Total Population')
```

```
legend('Output Neural Net',' Target','Location','Best')
```

```
figure,
```

```
plotregression(targets,outputs,'Regression')
```

The results of data experiments with the network model are presented in Table 2.

**Table 2.** Experiment Result

Model	Result		
	Iteration	Perform	Regresion
1	8	0.00094869	0.83259
2	7	0.00076587	0.83968
3	10	0.00046405	0.94414
4	7	0.0021427	0.3515
5	9	0.00081544	0.82761

#### 4. CONCLUSION

From the result of research can be concluded that Artificial neural network with training function Bayesian Regularization can do well forecasting. Best performance results are designated by neural net 3 model but require more repetition. The network 2 model is capable of converging with less iteration and less bad performance values. Speed and performance are necessities that can not be clearly compared. Both of these depend on user requirements.

**REFERENCES**

- [1] P. D. P. Silitonga, H. Himawan, and R. Damanik, "Forecasting acceptance of new students using double exponential smoothing method," *J. Crit. Rev.*, vol. 7, no. 1, pp. 300–305, 2020, doi: 10.31838/jcr.07.01.57.
- [2] H. Himawan and P. D. P. Silitonga, "Comparison of forecasting accuracy rate of exponential smoothing method on admission of new students," *J. Crit. Rev.*, vol. 7, no. 2, pp. 268–274, 2020, doi: 10.31838/jcr.07.02.50.
- [3] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 2014.
- [4] K. K. Aggarwal, Y. Singh, P. Chandra, and M. Puri, "Bayesian Regularization in a Neural Network Model to Estimate Lines of Code Using Function Points," *J. Comput. Sci.*, vol. 1, no. 4, pp. 505–509, 2005, doi: 10.3844/jcssp.2005.505.509.
- [5] S. Vhatkar and J. Dias, "Oral-Care Goods Sales Forecasting Using Artificial Neural Network Model," *Procedia Comput. Sci.*, vol. 79, pp. 238–243, 2016, doi: 10.1016/j.procs.2016.03.031.
- [6] K. Zhao and C. Wang, "Sales forecast in E-commerce using convolutional neural network," *arXiv*, 2017.
- [7] R. Aggarwal and R. Kumar, "Effect of Training functions of Artificial Neural Networks (ANN) on Time series Forecasting," *Int. J. Comput. Appl.*, vol. 109, no. 3, pp. 14–17, 2015, doi: 10.5120/19168-0634.
- [8] S. Shanmuganathan, "Artificial Neural Network Modelling: An Introduction," vol. 628, 2016, pp. 1–14.
- [9] J. Zou, Y. Han, and S.-S. So, "Overview of Artificial Neural Networks," *Methods Mol. Biol.*, vol. 458, pp. 14–22, 2009, doi: 10.1007/978-1-60327-101-1\_2.