

## Comparison of hybrid ANN models: A case study of instant noodle industry in Indonesia



Edy Fradinata <sup>1,2,\*</sup>, Sakesun Suthummanon <sup>2</sup>, Wannarat Sunthiamorntut <sup>3</sup>

<sup>1</sup>Industrial Engineering, Universitas Serambi Mekkah, Banda Aceh, Indonesia

<sup>2</sup>Industrial Engineering, Prince of Songkla University, Hatyai, Thailand

<sup>3</sup>Computer Engineering, Prince of Songkla University, Hatyai, Thailand

### ARTICLE INFO

#### Article history:

Received 8 April 2017

Received in revised form

23 June 2017

Accepted 29 June 2017

#### Keywords:

Hybrid method

Artificial Neural Networks

Analytic hierarchy process

Monte carlo

Geometric random

Forecasting model

### ABSTRACT

Artificial neural networks (ANNs) is the most stand popular practice to forecast the demand product since the other techniques still do not give the more accuracy. Furthermore, the hybrid method from ANN promises the best alternative to predict the customer demand. This paper proposes the hybrid model of ANN with the analytic hierarchy process (AHP), Monte Carlo (MC), and geometric random distribution to create new models to obtain the unusual methods in prediction. Those methods are substituted in the spaces of input weight and bias in the network. These hybrid methods are called  $AHP_{iw}ANN_b$  and  $MC_{iw}ANN_b$ . The hybrid technique of ANN has an approach of the time series-forecasting model. ANN is implemented in the testing case after the training process has run by the system, and process of validate is compared the testing from the training dataset. The Overall process is iterating the error to produce the mean squared error (MSE). The conclusions of this study, the hybrid ANN with AHP, MC and geometric random distributions show the good result of small MSE. More specifically, the hybrid AHPANN is better than hybrid MCANN.

© 2017 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Forecasting techniques are a fascinating subject to explore the data to predict the certain condition in the future. Some researchers compare forecast performance between traditional forecasting methods and data mining methods. The comparison of ANN method to some traditional forecasting methods such as multiple linear regressions, Holt's, Naïve and a moving average (MA) indicate the result that ANN is the best performance (Law, 2000). Other researchers have explored the comparisons of ANN with traditional forecasting methods such as arima and MA regression. They conclude that the ANN is better than the conventional methods (Kihoro et al., 2004; Fradinata et al., 2014). The methods of ANN, MA, exponential smoothing, and cubic regression develop and compare each model, it concludes from this study that ANN obtain the better result than the traditional methods. Lately, Artificial Neural Networks have developed particularly to hybrid with

other methods such as MC, AHP, Fuzzy and GA to predict more accuracy in many areas such as health, manufacturer, agriculture, management, and business. Consequently, it is vital to adopt a systematic approach to the development of this method to combine with some other methods to obtain the increase of performance. The determination of adequate model from the combination of the network could be implemented at four spaces; input weight, input layers, bias at the hidden layer, and the bias output layer. The researchers have been applied these combination methods to obtain the better performance compared to other methods in many cases.

Furthermore, the Fuzzy\_AHP is combined with ANN method incorporate with heart failure data. The data is used to input variable to ANN where it is previously normalized. Fuzzy\_AHP weights are used to change the input weights in neural networks in carrying out the training process. Furthermore, these two methods, Fuzzy\_ANN combination, and the original ANN are used to predict a patient's HF status, and their performances are observed from the results of output. The results of this research that the Fuzzy\_AHP\_ANN is better than the original ANN (Samuel et al., 2017).

\* Corresponding Author.

Email Address: [edinata69@gmail.com](mailto:edinata69@gmail.com) (E. Fradinata)

<https://doi.org/10.21833/ijaas.2017.08.004>

2313-626X/© 2017 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

AHP is combined with ANN, this study modifies AHP with ANN methods, the modifications are at the weight input layer and the weight output layer of ANN. He assumes that the bias is 1. This study is to measure the vendor selection. The results are meaningful in this study because it provides the hybrid to integrate AHP and ANNs techniques (Kumar and Roy, 2010).

Moreover, the AHP combination model with ANN has verified the outcomes of the fuzzy AHP model. The hybrid model is trained to predictable the most suitable machine for a good alternative to save a particular solution and give the effort to solve the problem for a new decision-making process. Furthermore, ANN is used to select the necessary machinery for the flexible manufacturing cell structure (FMC). The results for comparing the fuzzy AHP and ANN predict the model where it has proposed to decision support system. It could be chosen the most appropriate machine tool (Stam and Kuula, 1991).

ANN in this study cooperates with the weight of AHP to train to obtain the value of mean squared error on some network where it has been determined. The contribution of this research is to integrate the AHP with ANN method; it is called as the hybrid AHP-ANN method. It is a very attractive model in a hybrid model that involves both advantages methods (Tang et al., 2013).

The contribution of this paper is to estimate an instant noodle demand in hybrid ANN with AHP, MC and geometric random distribution. The novels of this study define clearly, where the geometric random normal distribution is selected from other random distributions to substitute to the bias in the network of ANN. Another contribution of this paper of the output is tested their consistency of robustness algorithm with the several of neurons at a hidden layer.

The last contribution, there is a variance test for various mean squared errors from the output. They are tested to see the consistency of variance with the eligible non-parametric Friedman test; this test is selected because some models are smaller than 30 data, it could be used in the linear and nonlinear dataset. Finally, this study hybrids ANN method with AHP, MC and geometric random distribution could be useful to encourage the academicians and researchers to develop with other methods in this area of knowledge.

The rest of this paper is structured as follows. Section two reviews work on the ANN method, AHP, MC methods, determinant the demand variables, and the measure the accuracy method. On section three, it sets some methodologies are used, and test of parameters of ANN, robustness test, at section four shows that the empirical result from the study. The discussion is in the fifth section where it discusses all the effect and advantage from the models. The final section is the conclusion of the paper where it reflects the performance of hybrid methods compares to the original ANN.

## 2. Methods and materials

### 2.1. The ANN method

The core component model of the neural network is perceptron. It considers the perceptron as an adaptive component behavior to resemble the neuron (Widrow and Hoff, 1960). A neuron, the initial construction is a lump for the information process to finalize the result could be seen in the Fig. 1a, a representative cell body of the nucleus, the signal from the external is feeding by dendrites to the cell body, and next signals are carrying signal cell to other agencies by the axons. It is interpreting that in an analog expression computational technology, as shown in Fig. 1b, the system is called a perceptron. It covers for element linear and a nonlinear. The input of signals,  $x_i$  are connected to the adaptable weighting components,  $w_i$  is the core part of the component, the signal output(s) entitled the bias. The result in the output of  $Y_o$ . Furthermore, the network of ANN works with the transfer function to make the iteration process to obtain the convergence the mean squared error on smoothly.

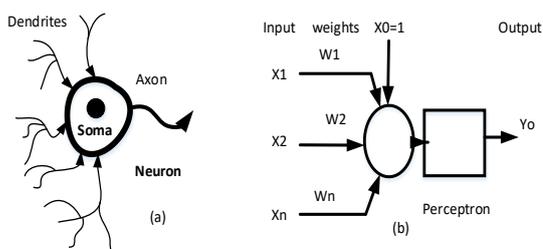


Fig. 1: Symbolic representations of neuron and perceptron

In the hidden layer computes  $y_i$  with the Eq. 1.

$$y_i = \sum_{i=1}^n I_i W_{i,j} \tag{1}$$

The following equation defines the output signal of  $y_o$  from the network (Eq. 2).

$$y_o = f(\sum_{i=1}^n w_i x_i + w_o) \tag{2}$$

Moreover, the bias follows the relationship is presented in the following Eq. 3.

$$w^T x + w_o \geq 0 \tag{3}$$

Then, the activation function cooperates to gather in the range of the hidden layer to drop in the certain range of activation function.

The perceptron will stimulate and yield the output signal when the condition of this is met. The study is for the selecting function the binary step of a log-sigmoid function in activation rule, (Fig. 2a) (Block, 1962; Hemming, 2003).

The common transfer function in neural network systems is sigmoid and tangen. They are illustrated in Fig. 2.

Fig. 2a and 2b show the form lines of the tanh and the sigmoid activation function. Both features that have some similarities with the function where the

line increases that the steepness approaches to point one. The mapping of the sigmoid function is  $x: \mathbb{R} \rightarrow (0,1)$  and in the case of a hyperbolic tangent, the mapping is a  $\text{tansig}(\tanh): \mathbb{R} \rightarrow (-1,1)$ .

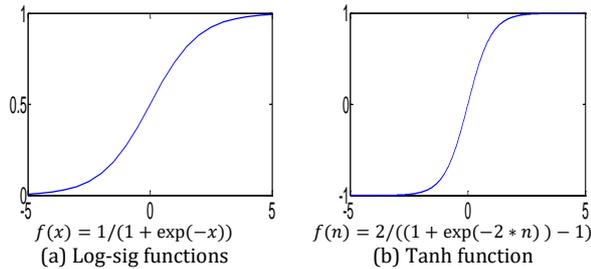


Fig. 2: (a) Logsig and (b) Tanh activation function

The continuous function is used from the activation function since this is necessary for the learning networks. The activation functions play a major role in determining the output of the functions. The learning processes continuously train on the set of the training parameter. The weights could be adjusted related its input data by the perceptron. Widrow and Hoff (1960) had made the delta rule for adjusting the perceptron weight; it is called a recursive gradient type of learning algorithm.

Finally,  $y_t$  is out from the output layer obtains the summation function in the following Eq. 4.

$$y_t = \sum_{j=1}^n f(y)_j W_{j,k} \tag{4}$$

As overall, the neural network has the input layers to receive the input variables and then process it into the hidden layers (Eq. 5).

$$w(k + 1) = w(k) + \eta \varepsilon(k) x(k) \tag{5}$$

The weight value  $w(k)$  is a represent  $\eta \varepsilon(k) x(k)$ . The bias creates the afterward weight value by an iteration process.

The approach of the neural network to the forecasting time series or autoregression model is defined as a general formula,  $Y_t = \mu + Q_1 \cdot Y_{t-1} + e$ , where  $\mu$  is constant,  $Q_1$  is the weight,  $Y_{t-1}$  is the input variable to the network of ANN, and  $e$  is the bias, then the drawing can be illustrated in Fig. 3. There are some parameters of the neural network; they are a training function, activation function, and neurons. A training function has the function to adjust or arrange the weight, and bias in the network to estimate parameter errors in the model.

The transfer function is used to cooperate with the weight and bias to iterate the form of the mean squared error which smaller than the initial error; the input suggested the same with the specific range of an activation function let it works more smoothly (Santosa, 2007).

Scaling is not strong needed (Karunanidhi et al., 1994) but the normalized is still suggested on performing to the certain range of the activation function.

In the traditional statistical models, the data usually need to transform to be normality. In a neural network, the probability distribution of input variables, it is not mandatory to be measured due to ANN can be work in some fields of data (Burke and Ignizio, 1992). Recently, it has been pointed out to produce the optimal condition of the MSE make the scaling on the relevance to transfer a function of min-max with the range of 0 to 0.9 into the network. There are some other normalize methods to variety the data drop into the certain range, such as min-max, scaling, mapstd and normalization methods (Santosa, 2007).

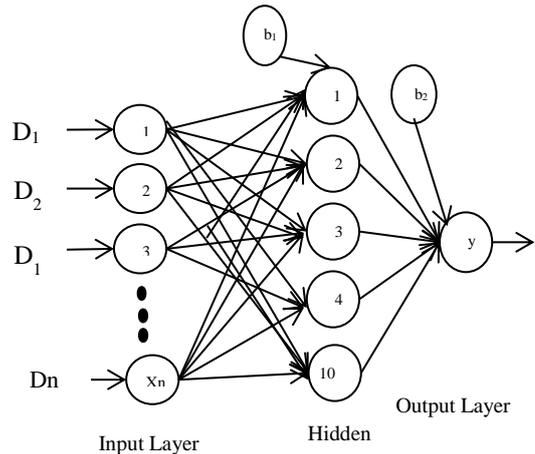


Fig. 3: Input-output nodes of ANN

However, if the data has already small enough drops into the range of transfer function, it does not need to make normalize it where it depends on the range of real dataset. The neurons in the network can be influenced the training process to obtain the over-fitting the computational time of training. Additionally, the internal training parameters such as the train param show, train param learning rate, epoch, time, goal, error, minimum gradient,  $\mu$  (mu) and validation check are co-operated to obtain the training process (Fradinata et al., 2014).

Cross-validation is one technique to assess the performance model in ANN modeling and contributing the good impact to separate data into testing; it could be used to generalize the ability to compare the training process to the testing case. The validating produces the new difference model to compare the testing model (Stone, 1974).

## 2.2. Analytical hierarchy process method

AHP is a technique to solve the problem from the hierarchy of the structure. It is used the multi-factor or multi-criteria transform into the hierarchy. According to Satty, The goal is a representation of an objective to solve the complex problem in a structure of multi-level, the second level factor is called the level of criteria, sub-criteria, and the last level is the alternatives.

The first step of AHP is determined the problem then make the hierarchical structure where it begins

from the primary objective. Furthermore defined the alternatives and made the pairwise comparison matrix, define the pairwise comparison, provide the comparison relative important table of Satty, determine the consistency index, and determine the consistency ratio (Satty, 2004, Ciptomulyono, 2008).

### 2.3. Monte carlo simulation method

MC simulation is the process of producing probability random independent where it forms from the specific random draw by a specified probabilistic model. This method is chosen because the MC simulation made the generated pool of data probability in the center of haul data. It could be specified by the multivariate normal distribution sampling or Latin Hypercube with one of the distributions. The random walk mostly uses the normal distribution. Excel or other software generates the random draw. This output should produce the useful data due to the data close to each other (Briggs et al., 2002).

### 2.4. Determinant of demand variable

Demand forecasting is the process of predicting demand need in the future from the sale or historical data. The statistical methods are utilized for extrapolating the future demand, where the varieties of methods are simple moving average, curve-fitting techniques, and time series analysis. The forecasting method depends on some pattern data factors (trends, seasonal, cyclic. Mostly, six determinants of demand variables should be influencing the market of the product; the population, the buyer, product price, the gross development product (GDP), promotion, and expected future price. Demand is the ability and willingness to buy specific quantities of goods in a given period at a particular price. Populations are some amount people in the country. The buyer is the amount of individuals or organization who buys the product. Price is the cost of the product. The GDP per capita is the monetary value of the finished goods and services of product in a country. Advertise involves the activities in promoting the product through the media. The expected future price is the price that a consumer is willing to obtain the certain price in the future.

### 2.5. Measure the forecast accuracy

The forecasting method develops the statistical method for generalizing historical data in the future. The method for forecasting future demand is based on various users and classification inputs. The different types of demand characteristics, such as unsmooth demand and seasonal demand are presented in statistical parameters. If the demand patterns show the significant result, the construct optimal forecasts will occur in future demand. The prediction of error in the case of tracking signals

where is calculated by the system would be determined the forecast and actual demand.

There are various measures of forecast errors. One of them is an *MSE*. The *MSE* is the square of quantity the unit term of an equal to the original value, which is the difference between the actual value and an estimated value; it can be defined in the calculating of the Eq. 6:

$$MSE = \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n}, \quad (6)$$

where:

$Y_t$  is an actual value for period  $t$ ,

$\hat{Y}_t$  is a period value of forecast  $t$ , and

$n$  is a period number

Cross-validation is producing an independent assessment of the good forecasting model to generalize the initial unseen data in statistical technique. The substance in cross-validation is the random split data into two parts, one training set and another is a test set. The neural network, the data part is used for training and estimating of the model's performance (He et al., 1997).

## 3. The purpose of hybrid model

The main objective of this study is to develop ANN methods from the ANN original; it is called the combination or hybrid method. From the hybrid method, it produced the new models. The methodology of this study can be seen in Fig. 4.

### 3.1. Data collection and identification variable

The data which were obtained from Indonesia Governments such as the Statistical Bureau of Indonesia, Badan Pusat Statistik Indonesia (BPS) (BPS, 2013) and Kementerian Perdagangan Indonesia (Trade Centre of Indonesia). The data had five variables from determinant of instant noodle demand dataset. The variables were identified with the Kolmogorov-Smirnov to check the normal distribution of data. Then, they were generated with a random normal distribution where they were spread on the normal distribution.

### 3.2. Variables selection

The data are generated by random normal distribution due to need longer dataset to fit well in training process then the original dataset. Other reason because each variables have not had the same years horizontally.

The variables are selected by the coefficient correlation significant or with the variance inflation factor (VIF) to choose the un-multicollinearity, the autocorrelation data in the same time is measured by the Durbin-Watson method. The non-significant variable relationship with the probability of  $\alpha > 0.05$  and the analysis of the obtain variable that the all variables are selected to be entrance variables. The selected variables are population, the price of

product, a gross development product (GDP), expected future price, and demand is used to normalize before it collaborates with the transfer function, weight, and bias in the network to find the optimal condition of the error. The demand is predicted from the other variables in the networks.

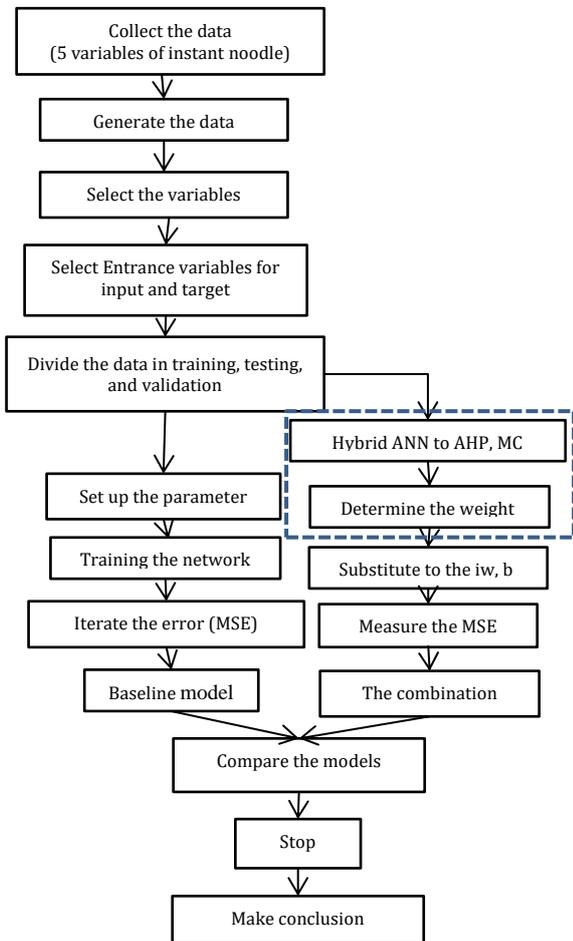


Fig. 4: Research methodology

The calculation of the expected future price per box is of \$6, \$6, and \$6. These data have the standard deviation of 0, 0.01, 0.31, and mean 0, 6.25, and 6.28. The generated probability number with standard deviation and mean have the function of probabilities value ( $f_p$ ) are  $2.48e^{-10}$ , 0.398. Then, the calculation as follow:

Expected #1. (Initial)

$$X_0 = 2.48e^{-10} / 2.48e^{-10} = 1$$

$$E(X_1) = 6 \times (1) = 6$$

Expected #2.

$$T_{tot} = 2.48e^{-10} + 0.398 = 0.398$$

$$(p)_1 = \frac{f_p}{T_{tot}} = \frac{(2.48e^{-10})}{(0.398)} = 6.23e^{-10}$$

$$(p)_2 = \frac{f_p}{T_{tot}} = \frac{(0.398)}{(0.398)} = 1$$

$$E(2) = \sum_{i=0}^n 71. (p_1) + 86. (p_2)$$

$$E(2) = 6 \times (6.23e^{-10}) + 6 \times (1) = 6$$

The calculation of the expected future price per box is of \$6, \$6, and \$6. These data have the standard deviation of 0, 0.01, 0.31, and mean 0, 6.25, and 6.28. The generated probability number with standard

deviation and mean have the function of probabilities value ( $f_p$ ) are  $2.48e^{-10}$ , 0.398.

Fig. 5 shows that the expected price values are under the actual data of price. It is because the customer tends to become a lower price than the original one

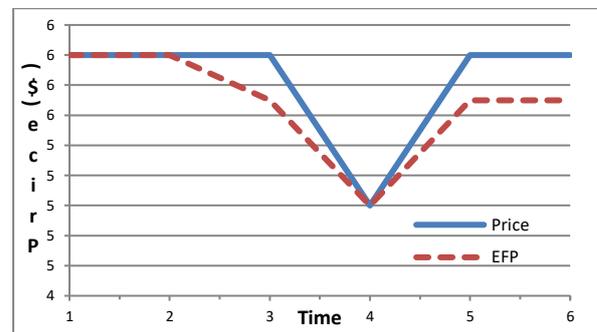


Fig. 5: Expected future instant noodle prices

### 3.3. Parameter selection

The selection of parameters such as layers, neuron, training function, and the transfer function is mandatory to select the relevance parameters when the data entrance to the network. These parameters kept the training process in the steady speed in the training process in the network of ANN. The selected parameters are applied to entrance variables to the neural network feed forward-back propagation. This study is used three layers; input layer, hidden layer, and an output layer with three layers, ten neurons, trainlm training function, and tansig as the transfer function. The selected parameters with the smallest MSE can be seen in Table 1.

This condition also is called the optimal parameters in the system of a neural network with the certain dataset.

### 3.4. Selected random distributions for input bias

Some random distributions are selected into ANN algorithm. The selected variables would be the smallest mean squared error. It is substituted into the bias in the network of the neural network. It is seen in Table 2.

Table 2 shows the geometric random distribution is chosen to be an entering variable to modify the bias in the neural network. The bias, AHP, and MC are combined to create new methods.

### 3.5. The AHP and monte carlo

The transformed some dataset from the input variables are converted to the rules of the system of Analytical Hierarchy Process software. The converted weights are used to substitute to the input weights of the network in the system. The notation of the input weight  $i_w$  is the weight into each variable of the network. The modified weight from Analytical Hierarchy Process replaces the default weight of a network in the hidden layer. Then, it cooperated to

the input layer in hidden layers to generate the better MSE than the initial case (Fig. 6).

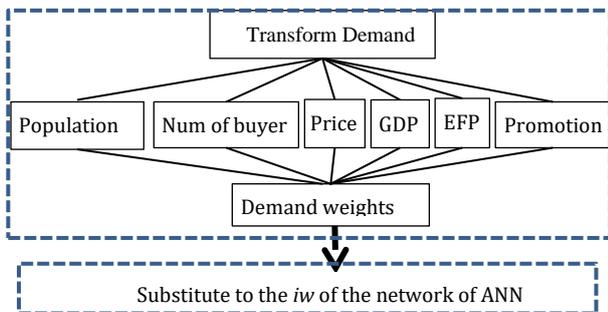
**Table 1:** Parameter selection of ANN

No	Layer	Neuron	Training function	Transfer function	MSE
1	3	N10	TrainLM	Tansig	2.65e <sup>-5</sup>
				Logsig	7.41e <sup>-4</sup>
				Purelin	2.53e <sup>-4</sup>
2			TrainGDM	Tansig	6.14e <sup>-3</sup>
				Logsig	2.62e <sup>-4</sup>
				Purelin	5.32e <sup>-4</sup>
3		N11	TrainLM	Tansig	9.22e <sup>-4</sup>
				Logsig	6.32e <sup>-3</sup>
				Purelin	7.77e <sup>-4</sup>
4			TrainGDM	Tansig	2.56e <sup>-3</sup>
				Logsig	4.87e <sup>-2</sup>
				Purelin	3.36e <sup>-4</sup>
5		N9	TrainLM	Tansig	5.69e <sup>-4</sup>
				Logsig	6.72e <sup>-4</sup>
				Purelin	8.29e <sup>-4</sup>
6	4	N10	TrainLM	Tansig	2.83e <sup>-4</sup>
				Logsig	1.38e <sup>-4</sup>
				Purelin	3.25e <sup>-3</sup>
7			TrainGDM	Tansig	4.62e <sup>-4</sup>
				Logsig	2.45e <sup>-4</sup>
				Purelin	4.37e <sup>-4</sup>
8		N11	TrainLM	Tansig	3.62e <sup>-4</sup>
				Logsig	2.47e <sup>-4</sup>
				Purelin	6.67e <sup>-3</sup>
9			TrainGDM	Tansig	7.43e <sup>-3</sup>
				Logsig	8.53e <sup>-4</sup>
				Purelin	3.74e <sup>-3</sup>
10		N9	TrainLM	Tansig	2.32e <sup>-3</sup>
				Logsig	7.30e <sup>-4</sup>
				Purelin	5.34e <sup>-4</sup>
11			TrainGDM	Tansig	4.81e <sup>-4</sup>
				Logsig	6.43e <sup>-4</sup>
				Purelin	3.35e <sup>-3</sup>
				Tansig	2.34e <sup>-3</sup>
				Logsig	4.25e <sup>-4</sup>
				Purelin	2.39e <sup>-4</sup>

Note: the value of MSEs is before post processing

**Table 2:** MSE from the variety of random distributions in the ANN network

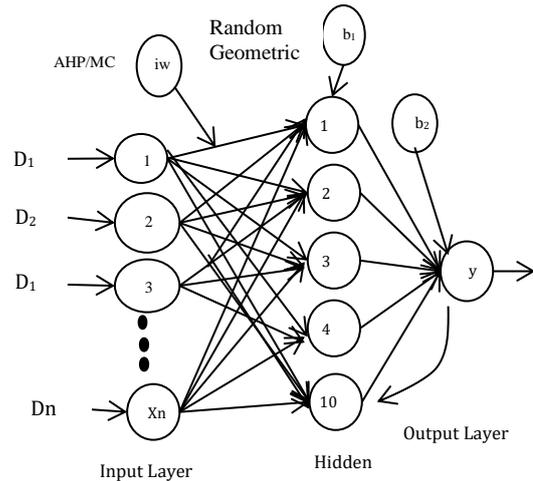
No	Random distribution	MSE
1	Binomial	3.03e <sup>-5</sup>
2	Chi-square	6.69e <sup>-4</sup>
3	Discrete U	1.14e <sup>-4</sup>
4	Exponential	6.39e <sup>-5</sup>
5	Gamma	4.07e <sup>-5</sup>
6	Geometric	1.53e <sup>-5</sup>
7	Normal	5.84e <sup>-5</sup>
8	Poisson	6.37e <sup>-5</sup>
9	Weibull	2.20e <sup>-4</sup>
10	Uniform	2.90e <sup>-4</sup>



**Fig. 6:** Transform input variables dataset to the AHP system

Meanwhile, the output of Monte Carlo (MC) simulation resulted from the normal random

generation to form the random draw. Then some samples of data are taken from it to substitute the bias at the part of network. It associates to input variables, input weights of the network before substitute to the transfer function in the hidden layer. This method is chosen because the MC simulation made the generated pool of data probability in the center of data. This output should produce the good data because of data close to each other. The data is iterated to obtain the smaller error than the initial error in the hidden layers in the training process. The output of AHP and MC substituted to the input weight and input bias in the ANN system; the figure can be seen in Fig. 7.



**Fig. 7:** The Purposed hybrid network of ANN

The  $x_1, x_2, \dots, x_n$  are scalar demand. Input,  $w_1$  is the network weight,  $b_1$  is the bias in the network of ANN. The scalar input transmitted through a connection that multiplies its strength by the scalar weight with the transfer function of in the hidden layer. The weight update equation for the classical Newton algorithm is shown by (Nolfi and Parisi, 1996) (Eq. 7).

$$w_{n+1} = w_n - H^{-1} \nabla_{w_n} E \tag{7}$$

Eq. 7 shows that  $H^{-1}$  is the inverse of the Hessian matrix where the iteration process needs space to give to  $k^2$ , the notation of  $k$  is the free parameter number (Kwok and Yeung, 1997). The hidden layers continue the process to iterate before sending the process to the output layers. The modification of the weight could be substituted for the input layer.

**3.6. The algorithm of ANN is as follow**

The algorithm feedforward backpropagation follows the phases as follow:

*Phase-1. Input*

1. Each input unit receives the signal and forwards it to the hidden layer.  $z_j, (j=1,2,\dots)$ .

$$z_{net_i} = v_{jo} + \sum_{i=1}^n x_i v_{ji} \tag{6}$$

$$z_j = f(z_{net_i}) = \frac{1}{1+e^{-z_{net_j}}} \quad (7)$$

2. Count all outputs in hidden  $y_k, (k=1,2,\dots,m)$

$$y_{net_k} = w_{k0} + \sum_{j=1}^p z_j w_{kj} \quad (8)$$

$$y_k = f(y_{net_k}) = \frac{1}{1+e^{-y_{net_k}}} \quad (9)$$

3. Count all network output in the  $y_k(k=1,2,\dots,m)$

Phase-2. Back propagation

4. Count the factor  $\delta$  arithmetic unit output by an error in each unit of output  $y_k(k=1,2,\dots,m)$ .

$$\delta_k = (t_k - y_k) f'(y_{net_k}) = (t_k - y_k) y_k (1 - y_k)$$

5. The  $\delta_k$  is a unit of error that will be used in a layer below the weight change.

$$\Delta w_{ij} = \alpha d_k z_j ; k = 1, 2, \dots, m; j = 0, 1, \dots, p$$

6. Calculate the weight change rate  $x$  with the accelerated pace of  $z_j (j = 1, 2, \dots, p)$

$$\delta_{net_j} = \sum_{k=1}^m \delta_k w_{kj}$$

7. The factor  $\delta$  in the hidden layer:

$$\delta_j = \delta_{net_j} f'(z_{net_j}) = \delta_{net_j} z_j (1 - z_j)$$

net.iw{1,1};%=AHPweight

net.b{1};%=bias1

net.lw{2,1};%=AHPw2

net.b{2};%=bias2

8. Calculate the change in weight,  $v_{ji}$ .

$$\Delta v_{ji} = \alpha \delta_j x_i ; j=1,2,\dots,p ; i=0,1,\dots,n.$$

Phase 3. Change the weight

9. Calculate all weight changes.

$$w_{kj}(new) = w_{kj}(old) + \Delta w_{kj}, (k = 1, 2, \dots, m)$$

and  $j=(0,1,\dots,p)$

10. The weight change to the hidden layer is leading to output.

$$v_{ji}(new) = v_{ji}(old) + \Delta w_{kj}, (j = 1, 2, \dots, m)$$

and  $i=(0,1,\dots,p)$

### 3.7. Robustness test

The algorithm of robustness work with the statistical mean squared error from varies the neurons. The neurons are combined from 10 to 80 in the hidden layer. Each neuron iterated to obtain the mean squared error for its performance and It can be seen in Fig. 8.

The result indicated that the MSE is closed to the average data for each point. It means the algorithm is robust.

### 4. Empirical result

The result of forecasting model shows in Fig. 9 where the prediction points are from 80 to 100 axis

points after the training process. The testing and validating process work randomly from the training data to predict 20 points. The testing and validating are developed to obtain the accuracy of prediction.

This condition illustrated that it represents the testing and validation data have the correlation R-value are in good condition.

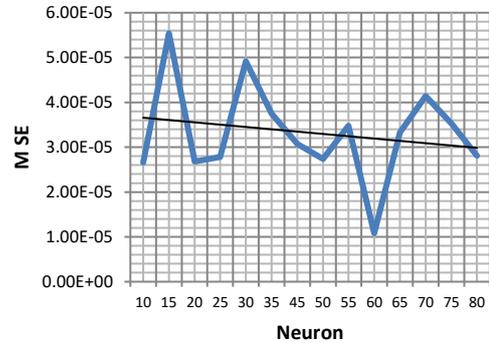


Fig. 8: Robustness test of algorithm

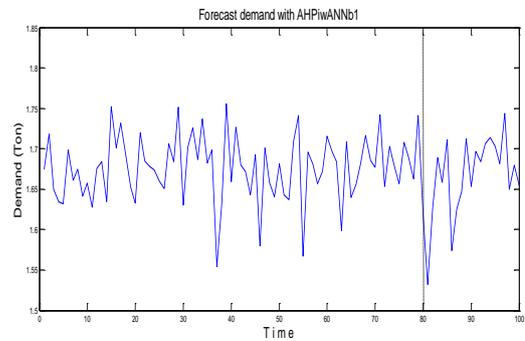


Fig. 9: Forecast demand of instant noodle industry

The lines data represented the overlap lines from the three processes. It means that the validation has a good performance because the process of the best-fit regression line between output and target relationship closed to 1.000. It concludes that the perfect prediction in the process of forecasting models from the process. The prediction error process worked on the global minimal. The optimal condition of validation is chosen on the value of 1.000.

This study is used in seven combination methods where they are created from the original ANN with the AHP, MC, and random geometric distribution. The original ANN is a baseline comparison from the other combination methods. These new varied methods are as follow: The order of  $i_w$  is the input weight, and  $b$  is the bias. These notations show the varied of the combination methods within the neural network. The abbreviation of hybrid ANNs are illustrated as follow:

AHP<sub>iw</sub> AHP weight substitutes in a part of input weight of ANN.

MC<sub>iw</sub> MC weight substitutes in part of input weight of ANN.

ANN<sub>b1</sub> Geometric random distribution works at the bias hidden layers of ANN.

ANN<sub>b2</sub> Use 1 as a bias at output layer of ANN.

The mean squared error results from combination methods are shown in Table 3.

**Table 3:** The MSE post process of hybrid models for instant noodle industry

Model	MSE
AHP <sub>iw</sub> ANN <sub>b1</sub>	3.67e <sup>+6</sup>
MC <sub>iw</sub> ANN <sub>b1</sub>	3.41e <sup>+8</sup>
AHP <sub>iw</sub> ANN	5.58e <sup>+10</sup>
MC <sub>iw</sub> ANN	6.10e <sup>+10</sup>
AHP <sub>b1</sub> ANN	2.74e <sup>+10</sup>
MC <sub>b1</sub> ANN	4.27e <sup>+10</sup>
ANN <sub>b1</sub>	5.48e <sup>+10</sup>
ANN	7.28e <sup>+10</sup>

The MSEs tend to increase from the AHP<sub>iw</sub>ANN<sub>b1</sub> to the original ANN. Mean that the AHP<sub>iw</sub>ANN<sub>b1</sub> has the smallest value of MSE.

### 5. Discussion

Neural networks are commonly used to forecast demand variables. This study obtained the determinant of demand instant noodle variables in Indonesia. The data was utilized as input variables to the network, AHP, MC was entranced as input weights, and the geometric random substituted the bias in the hidden layer.

These results can be explained by divided into three parts of the process: *first* is the pre-processing. This process referred to supervise the neural network that is normalized the input variables from determinant of demand dataset. It was normalized with the mapstd where this function drops the data in the range between 0 to 1. The autocorrelation was measured by the Durbin-Watson, and the correlation relation table measured the correlation among the variables independent with the significant correlation. The result obtains the correlation higher number than the significant, means that the correlation belongs the independent variables were weak and independently. *Second*, the chosen variables were transferred to the transfer function that had relevance range to able to process in the system of the neural network. The system worked with the optimal parameters for the detection of trends in the learning process. The parameters were determined to identify the distribution to make the relevance patterns.

The training process learned the pattern of data characteristic of data to iterate the error that works in the range of the transfer function capability at the upper bound and the lower bound to produce the mean squared error. At the end of the process, the data to reverse back to the real data. The advantage of ANN is the ability to adapt the condition through the retraining process when the input data entrance to the network. This study is set the training process for 80 data. The 80 data elaborate together with the weight of AHP or MC, and with the bias of geometric random distribution. Due to input variables, the weight, and the geometric random variable were small enough to come to be the smaller iteration than the initial in each iteration process. Hybrid models were obtained of input variables from AHP,

MC and geometric random in the network. The AHP combination ANN is better than the MC combination ANN. It is because the transform data to AHP produced the smaller weight due to the AHP weight which validates the ratio consistency index smaller than 0.1. This condition contributed to the process of random work small in the network to iterate the mean squared error compared to the MC. The MC normal distribution is selected and has the random number up to 0.2. The modification of ANN substituted on parts of input weight, an input layer, and bias. The AHP and MC entranced in the input weight and the input layer while the geometric random distribution entranced in the bias part. The combination methods were called AHP<sub>iw,il</sub>ANN<sub>bn</sub> and MC<sub>iw,il</sub>ANN<sub>bn</sub> combination. Subscript notations showed that the *b* (input bias), *iw* (input weight), and *il* (input layer). The various combination methods from instant noodle industries were AHP<sub>iw</sub>ANN<sub>b1</sub>, MC<sub>iw</sub>ANN<sub>b1</sub>, AHP<sub>iw</sub>ANN, MC<sub>iw</sub>ANN, AHP<sub>b1</sub>ANN, MC<sub>b1</sub>ANN, and ANN<sub>b1</sub>. The AHP<sub>iw</sub>ANN<sub>b1</sub> was better performance MSE and BWE than the MC<sub>iw</sub>ANN<sub>b1</sub>. The AHP<sub>iw</sub>ANN was better than the MC<sub>iw</sub>ANN. The AHP<sub>b1</sub>ANN was better than MC<sub>b1</sub>ANN. Furthermore, the ANN<sub>b1</sub> was better than original ANN. The AHP<sub>iw</sub>ANN<sub>b1</sub> was better than MC<sub>iw</sub>ANN<sub>b1</sub> because the value of AHP as input weight has smaller than the MC input weighed. These because the initial iteration on the random values of AHP and geometric random as back propagation algorithm trained a bias with the AHP value is quite close to zero. It means that the result of an iteration process from these two factors collaborate should have the small result of prediction error. Then, the testing process from this hybrid method was tested after the training process, and the validating process worked to compare the real data to the predicted dataset. The error function from this hybrid minimized the mean squared error. *Third*, the post processing data of forecasting demand dataset. It was the retransferred to the original dataset after the process obtains the smallest mean squared error. The combination ANN with AHP was better than ANN with MC. *At last*, there were two methods to examine the results, *first* the test of the algorithm used to check the robustness of prediction accuracy Fig. 7. *Second*, the variances were tested with the Friedman test. Moreover, at the end of this study; we measured the variance performance between the models with Friedman test. The Friedman test can be used in linear and non-linear data. The other reasons because the numbers of samples are less than 30 and the variance of data are not spread on stationery. The Friedman test measures the consistency of variances among the models where the SPSS software is used to produce the output. The process of data measurement was separated into two groups, MC and AHP groups. The result can be seen in Table 4.

As a result, the computed p-value is 0.083 where it is higher than the significance level  $\alpha > 0.05$ ; it means that the performance of the MSE samples

from different hybrid models AHP and MC had no variance among instant noodle dataset.

**Table 4:** Friedman statistics test the MSE

N	4
Chi-square	3.000
Df	1
Asymp.sig.	0.083
a.Friedman Test	

## 6. Conclusion

The neural network is a popular and widely used method to estimate the certain condition in the future. This study proposes the supervised back-propagation hybrid of ANN from the AHP, MC and the geometric random distribution to obtain the better performance than the original ANN. The modified methods between original ANN and additional methods are called the combination (hybrid) method of AHP<sub>iw</sub>ANN<sub>b1</sub>, and MC<sub>iw</sub>ANN<sub>b1</sub>. The MSE is used to measure the performance of the model. The results from the hybrid methods are better performance than the original one. The smallest of MSE occurs at the AHP<sub>iw</sub>ANN<sub>b1</sub> with the value of  $3.67e^{+6}$ .

The comparison model of combination AHP is better than combination MC. The other hybrid models such as AHP<sub>iw</sub>ANN, MC<sub>iw</sub>ANN, AHP<sub>b1</sub>ANN, MC<sub>b1</sub>ANN, and ANN<sub>b1</sub> also perform the excellent promising to be forecasting models compared to original ANN. The combination methods have better performance than the original ANN method because of some reasons; the bias runs into the smallest mean squared error with the geometric probability distribution. The input weight works with the AHP and MC methods where they have the advantage to make the smallest result where they quite close to zero with the random normal characteristic. It is helpful when the training process estimates the model. Some methods are used to check the algorithm back propagation neural networks to test the process are quite good. The robustness test shows that the parameters work in optimal conditions from the algorithm. The other test methods are Friedman test to test the variance among the models.

## 7. Future work

The suggestion for the future work is trying to develop more models to hybrid with other original methods with data mining and artificial intelligent methods.

## Acknowledgement

The authors would like to thank Kemristek Dikti of Indonesia who has given the chanced for the scholarship program, to the Prince of Songkla University, Hatyai, Thailand for their supporting this research and Thomas Dryden from England (UK) as the proofreader.

## List of symbols

$w_i$	Network weight
$Z_j$	Hidden layer
$y_k$	Outputs in hidden
$\delta_k$	Unit of error used in the layer
$\Delta w_{ij}$	The weight change
$v_{ji}$	Calculate the change in weight
$w_{kj}(new)$	Calculate all weight changes new
$v_{ji}(new)$	Weight change to the hidden layer
$\mu$	Mean
$D_t$	Signal demand random normal
$D_{t+1}$	Demand during period
$e_n$	Error forecast
$Y_t$	Actual value for period
$n$	Period number
$y_0$	Output signal
$w_i$	Initial weight
$x_i$	Initial variable
$w_o$	The weight
$f(x)$	Function of x
$w^T$	Transform weight
$i_w$	Input weight
$i_l$	Input layer
$b_{1,2}$	Bias 1 and 2.
$E(x)$	Expectation
$F(x)$	Function of probability $X_i$
$f_p$	Probability value
$d$	Durbin-Watson value
$X_1, X_2, \dots, X_n$	Scalar demand

## References

- Block HD (1962). The perceptron: A model for brain functioning. i. Reviews of Modern Physics, 34(1): 123-134
- BPS (2013). Proyeksi Penduduk Indonesia. Badan Pusat Statistik Indonesia. Jakarta, Indonesia. Available online at: <https://www.bps.go.id/>
- Briggs AH, Goeree R, Blackhouse G, and O'Brien BJ (2002). Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. Medical Decision Making, 22(4): 290-308.
- Burke LI and Ignizio JP (1992). Neural networks and operations research: an overview. Computers and Operations Research, 19(3-4): 179-189.
- Ciptomulyono U (2008). Fuzzy goal programming approach for deriving priority weights in the analytical hierarchy process (AHP) method. Journal of Applied Sciences Research, 4(2): 171-177.
- Fradinata E, Suthummanon S, Sirivongpaisal N, and Suntiarnornthuthq W (2014). ANN, ARIMA and MA timeseries model for forecasting in cement manufacturing industry: Case study at lafarge cement Indonesia—Aceh. In the International Conference of Advanced Informatics: Concept, Theory and Application, IEEE, Bandung, Indonesia: 39-44. <https://doi.org/10.1109/ICAICTA.2014.7005912>
- He H, Wang J, Graco W, and Hawkins S (1997). Application of neural networks to detection of medical fraud. Expert Systems with Applications, 13(4): 329-336.
- Hemming C (2003). Using neural networks in linguistic resources. Department of Languages, University College of Skövde, Swedish National Graduate School of Language Technology, Skövde, Sweden. Available online at: <http://hemming.se/gsl/LingRes/nnHmg.pdf>

- Karunanidhi K, Natarajan N, and Thangaraju P (1994). Genetic and Non-Genetic variation in birth placental weight and number of cotyledons in Mecheri and its Dorset cross sheep. *Cheiron*, 23: 255-260.
- Kihoro J, Otieno R, and Wafula C (2004). Seasonal time series forecasting: A comparative study of ARIMA and ANN models. *African Journal of Science and Technology*, 5(2): 41-49
- Kumar J and Roy N (2010). A hybrid method for vendor selection using neural network. *International Journal of Computer Applications*, 11(12): 35-40.
- Kwok TY and Yeung DY (1997). Constructive algorithms for structure learning in feedforward neural networks for regression problems. *IEEE Transactions on Neural Networks*, 8(3): 630-645.
- Law R (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management*, 21(4): 331-340.
- Nolfi S and Parisi D (1996). Learning to adapt to changing environments in evolving neural networks. *Adaptive Behavior*, 5(1): 75-98.
- Samuel OW, Asogbon GM, Sangaiah AK, Fang P, and Li G (2017). An integrated decision support system based on ANN and Fuzzy\_AHP for heart failure risk prediction. *Expert Systems with Applications*, 68: 163-172.
- Santosa B (2007). Data mining teknik pemanfaatan data untuk keperluan bisnis. Graha Ilmu, Yogyakarta, Indonesia.
- Satty TL (2004). Decision making—the analytic hierarchy and network processes (AHP/ANP). *Journal of systems science and systems engineering*, 13(1): 1-35.
- Stam A and Kuula M (1991). Selecting a flexible manufacturing system using multiple criteria analysis. *The International Journal of Production Research*, 29(4): 803-820.
- Stone M (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)*, 36(2): 111-147.
- Tang SH, Hakim N, Khaksar W, Ariffin MKA, Sulaiman S, and Pah PS (2013). A hybrid method using analytic hierarchical process and artificial neural network for supplier selection. *International Journal of Innovation, Management and Technology*, 4(1): 109-111
- Widrow B and Hoff ME (1960). Adaptive switching circuits. In the *IRE WESCON Convention Record*, 4(1): 96-104.