The hybrid-model architectural modelling based on ARIMA-BPNN methods for building materials demands forecasting

Cynthia Hayat^{1,*}, Iwan Aang Soenandi²

¹Dept of Information System, Krida Wacana Christian University, 11440 Jakarta, Indonesia

²Dept of Industrial Engineering, Krida Wacana Christian University, 11440 Jakarta, Indonesia

Abstract. The development of hybrid architectural model is made to facilitate the decision making in determining the demand for building materials procurement. The ARIMA time series and non-linear BPNN models are selected considered that they are able to have a high degree of accuracy of the generated output. The data used were the secondary data collected in the period of February 2015-October 2016 which consisted of month of sales period, product prices, sales history (per type of building materials), estimated number of renovation projects, estimated number of new construction projects, and number of competitors. This research was conducted through 2 stages, they were; the processing of time series using ARIMA through three basic steps, namely identification, assessment and testing, and diagnostic examination; and the BPNN processing through data training and data testing stages. The produced hybrid architectural model had 99% accuracy with an MSE of 0.00099926 on epoch 975 and training period of 00:00:01. The Regression results showed that the produced model has a high degree of accuracy in generating the output of building materials demand forecasting.

1 Introduction

In improving the efficiency of the supply chain system, there is a need of demand forecasting because each stakeholder involved makes a demand order in response to a demand signal [1]. A good demand forecasting is which can have accurate estimation so as to facilitate the operations of the company such as in planning the material needs and inventory management [2]. A building materials company needs to consider both internal and external aspects that may have some effect on its operations. The most often faced internal aspects are the limited capital or financial flows as well as the space (business place) for storage and mobility of goods. Meanwhile, from the external aspects, the most influencing aspect is the number of material types which consists of various types of goods with various types, brands, and prices, the level of difficulty of the material source, the number of competitors, as well as market

* Corresponding author: cynthia.hayat@ukrida.ac.id

conditions or consumer consideration factors in using building materials.

Time series forecasting is a qualitative method used for conducting a projection on future value through a certain time series so as to facilitate the decision making [3]. The application of the time series with the Auto-regressive Integrated Moving Average (ARIMA) model belongs to one of the most important research areas and is one of the most frequently used time series models. This is because the ARIMA model is quite flexible and able to represent several time series in a certain period, but has limits in making a forecast in a linear form. Meanwhile, in real world systems, there are many non-linear problems which are complex, so that the forecasting is not enough [4]. In the research conducted by Khasei and Bijari, it is said that the parameters used by the integration model of linear ARIMA and non-linear ANN in a time series can improve the performance and the desired prediction and used as the prediction value and the residual noise which will be used by the ANN model [5]. Wei used the ANFIS method to make a forecast by using fuzzy inference in a time series. Li Xiong and Yue Lu said that the use of both linear and non-linear models can improve the accuracy of forecasting results [6].

Thus, this study proposed a Hybrid approach by combining the linear ARIMA model with the non-linear Backpropagation ANN (BPNN) model. Using a hybrid model is estimated to be able to accommodate the real-time time series in which it is not completely linear nor non-linear [7][8]. The research question developed is "How to develop architectural model with the hybrid ARIMA-BPNN model to forecast the inventory of types of building materials?". With this hybrid model, it is expected to improve the accuracy of the expected demand forecasting. The purpose of this research is to generate a hybrid model from linear model of ARIMA and nonlinear model of BPNN in order to forecast the inventory of material type of building materials with high accuracy.

The results of several researches in time series forecasts show that the limitations on traditional statistical models that can only accommodate the linear data, so that a model is developed to be able to take into account certain nonlinear patterns observed in real problems in a time series. Nonparametric models also cannot significantly improve the accuracy. Therefore, hybrid model is one the models that can be used to forecast time series. The novelty of this research was on the research object by accurately forecasting the demand for building materials. In addition, the non-linear Backpropagation ANN method combined with linear ARIMA method is still rarely used.

2 Fundamental Hybrid ARIMA-BPNN Method

2.1 ARIMA-BPNN Hybrid Model

The ARIMA-ANN Hybrid Methodology was firstly introduced by Zhang in 2003. Zhang assumed that the time series data can be categorized as linear and non-linear data [9]. In 2011, Khasei and Bijari proposed a different approach where ARIMA and ANN should not to be adaptive where time series is of which included in both linear and non-linear components [10].

The ARIMA model consists of three basic steps, namely identification, assessment and testing, and diagnostic examination. Furthermore, ARIMA model can be used for forecasting if the model obtained is sufficient.

The ANN network architecture used in this research was the *multi layer network* with backpropagation algorithm. In this network, other than the input and output units, there can be one or more other units called the *hidden layer* which is determined in the training data. The multilayer neural network can solve the more complex problems compared to the single layer one, although it sometimes requires the more complex and

longer training process. The design of artificial neural network architecture was made by determining the numbers of input layer neurons, hidden layer and output layer neurons

The Hybrid model developed in this paper is attached in Figure 1 below:

Fig 1. The hybrid ARIMA-BPNN model for forecasting.

Figure 1 above describes the development of hybrid model of ARIMA-BPNN to forecast the inventory of building materials. Broadly speaking, the model development is conducted through 2 stages namely ARIMA Model process then BPNN Model. The BPNN Model consists of input layer, hidden layer, and output layer. The analysis result of ARIMA model is used as input layer on BPNN Model. The variable in the input layer is the period of month in selling history, the amount of historical selling, the price of the product, the approximation of project renovation, the approximation of building project, and the amount of competitor. While the output layer produces the forecasting of required building material supply.

2.2 Procedure of The Proposed ARIMA-BPNN Hybrid Model

The procedures of the proposed BPNN model development are as follows :

Step 1 : Pre-processing data was conducted by collecting the necessary data for conducting the data training. The data used were the secondary data collected from February 2015 to October

2016, consisted of several variables; month of sales period, product price, sales history (per type of building materials), estimated number of renovation projects, estimated number of new construction projects, and

Step 2 : Normalizing the input (x_1, x_2, x_3) and output (y) data using the formula

$$X = \frac{0.9(X-b)}{(a-b)} + 0.1 \quad [1]$$

- Step 3 : Conducting identification stages on ARIMA model. The identification stages were conducted based on the actual data plot that has been stationary. If it had not been stationary, the data must be made stationery first by using the formula $X' = X^{-2}$, or multiplied by $1 / \lambda$. Next, determining the possible combination of ARIMA model. From the autocorrelation plot, it was by determining the order of MA (q), from the partial autocorrelation plot, it was by determining the the order of AR (p).
- Step 4 : Performing the assessment stage on the ARIMA model. In this stage, there were estimation and testing of possible ARIMA model and the selection of the best model conducted.
- Step 5 : Performing diagnostic examination stage on ARIMA model. This stage was conducted by determining the equation and the forecasting value of the beast ARIMA model.
- Step 6 : Determining the number of nodes, activation functions and parameters used in the hybrid model. The activation functions served to limit the value of the desired forecasting output. The activation functions used were logsig-purelin, mean-square error (MSE) of 1.10^{-3} , while the learning rate used was 0.1 with a momentum constant (mc) of 0.95. Whereas, the iteration of training was conducted for 1000 epochs in order to get a good result.
- Step 7 : Conducting data training. Data training was conducted to validate the network model with multiple times of trial and error in order to get the best network by determining the number of neurons until obtaining the low error value.
- Step 8 : Testing the parameters of BPNN model. The weight of ANN generated from the training phase would be implemented into the testing

2.3 Data

The data used were the secondary data as in table 1 below. The data used were the data from the the period of February 2015 to October 2016, consisted of several variables: Month of Sales Period, Product Price, Sales History (per type of building materials), Estimated number of Renovation Projects, Estimated number of New Construction Projects, and Number of Competitors. In the determination of the reverse network, the data were divided into two, they are; training data and testing data. From 22 data, it was determined that there were 15 data for training and 7 data for testing.

Table 1. Secondary data and normalization of data input and target.

X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	Y
Feb 2015	80.000	25	8	1	18	50
Mar 2015	80.000	42	14	2	18	20
Apr 2015	85.000	128	2	5	18	100
May 2015	85.000	28	12	1	18	50
Jun 2015	81.000	53	18	2	18	50
Jul 2015	81.000	92	11	3	18	50
Aug 2015	81.000	30	3	1	18	50
Sep 2015	81.000	42	7	1	17	50
Oct 2015	81.000	5	3	0	17	50
Nov 2015	82.000	51	9	1	17	50
Dec 2015	82.000	236	3	2	18	200
Jan 2016	82.000	53	4	1	18	100
Feb 2016	82.000	92	8	1	18	50
Mar 2016	85.000	12	9	0	18	50
Apr 2016	85.000	53	5	1	18	50
May 2016	85.000	64	6	2	18	50
Jun 2016	85.000	212	6	4	19	200
Jul 2016	83.000	128	9	1	19	100
Aug 2016	83.000	53	3	1	19	50
Sep 2016	83.000	16	2	0	19	50
Oct 2016	83.000	131	5	3	19	100
Nov 2016	83.000	27	8	1	19	50

	X ₂	X ₃	X ₄	X ₅	X ₆	Y
0,17	0,10	0,17	0,40	0,26	0,50	0,23
0,25	0,10	0,23	0,70	0,42	0,90	0,10
0,32	0,90	0,53	0,10	0,90	0,90	0,46
0,39	0,90	0,18	0,60	0,26	0,90	0,23
0,46	0,26	0,27	0,90	0,42	0,90	0,23
0,54	0,26	0,40	0,55	0,58	0,90	0,23
0,61	0,26	0,19	0,15	0,26	0,90	0,23
0,68	0,26	0,23	0,35	0,26	0,85	0,23
0,75	0,26	0,10	0,15	0,10	0,85	0,23
0,83	0,42	0,26	0,45	0,26	0,85	0,23
0,90	0,42	0,90	0,15	0,42	0,90	0,90
0,10	0,42	0,27	0,20	0,26	0,90	0,46
0,17	0,42	0,40	0,40	0,26	0,90	0,23
0,25	0,90	0,12	0,45	0,10	0,90	0,23
0,32	0,90	0,27	0,25	0,26	0,90	0,23
0,39	0,90	0,30	0,30	0,42	0,90	0,23
0,46	0,90	0,82	0,30	0,74	0,95	0,90
0,54	0,58	0,53	0,45	0,26	0,95	0,46
0,61	0,58	0,27	0,15	0,26	0,95	0,23
0,68	0,58	0,14	0,10	0,10	0,95	0,23
0,75	0,58	0,54	0,25	0,58	0,95	0,46
0,83	0,58	0,18	0,40	0,26	0,95	0,23

It should be noted that most periodic sequences are non- stationary and that AR and MA aspects of the ARIMA model are only related to stationary periodic sequences. Stationarity means having no growth or decline in the data. The data must roughly be horizontal along the time axis. In other words, the fluctuation of the data around a constant average value, is independent of the time and variance of the fluctuation which essentially remains constant at all times.

A non-stationary time series must be converted into stationary data by performing *differencing*. *Differencing* is to calculate the change or the difference in the value of observation. The difference value obtained is checked further whether it is stationary or not. If it is not stationary, then there will be another *differencing* performed. If the variance is not stationary, then there will be a transformation of algorithm.

3 Result of ARIMA-BPNN Hybrid Model

3.1 Result of ARIMA Process

It should be noted that most periodic sequences are non-stationary and that AR and MA aspects of the ARIMA model are only related to stationary periodic sequences. Stationarity means having no growth or decline in the data. A non-stationary time series must be converted into stationary data by performing *differencing*.

A scaled sequence is said to be stationary or showing a random error if the autocorrelation coefficient for all lags, the numbers shown at each interval, are

statistically not different from zero or different from zero for just several lags ahead. An autocorrelation coefficient is said to be not different from zero if it is in the interval. There was a Lambda (Rounded Value) = -0.5 obtained in the stationarity step in Fig 2 and Fig 3 below.

Fig 2. The initial stationarity plot graphic.

Fig 3. The stationarity graphic after the transformation of initial data.

Figure 2 illustrates the result of stationary / random error correlation coefficient on the range obtained from Lambda results (rounded value = -0.5) Figure 3 illustrates the results of stationary / random error correlation coefficients after initial data transformation.

A data is said to be stationary to the variance and the mean, if the *rounded value* or $\lambda = 1$, then the existing data must be transformed firstly in order to have the Lambda value optimal by using the following formula: $X' = X^{-2}$, or multiplied by $1 / \lambda$, [2].

The result obtained from the initial data had not been stationary yet, therefore, there was a transformation conducted on the initial data. In Minitab software, the forecasting process was able to immediately conduct after getting the significance from ARIMA model. This process was conducted by trial and error method on its tentative model (1-1-0 or 0-1-1 or 1-1-1), in order to find the best model to be used. According to the significance obtained at the analysis stage, we used the orders of $p = 1$, $d = 1$ and $q = 1$, and the results obtained were in accordance with figure 4 below:

Fig 4. The training processes on ARIMA model.

3.2 Result of BPNN Process

Table 2 below describes the results of the BPNN process at the training data stage. The results displayed are based on the number of hidden parameters of the layer and the number of neurons used, the learning rate magnitude and momentum constraint that has been determined from the beginning, the type of activation function used, and the mean square error (MSE) generated.

Table 2. Training with different numbers of neurons to find the MSE.

No	Hidden layer	Neuron	Learning rate	Momentum constraint	Fungsi aktivasi	MSE
1	1	5-1	0.1	0.95	logsig-purelin	0.00403
2	1	6-1	0.1	0.95	logsig-purelin	0.00160
3	1	7-1	0.1	0.95	logsig-purelin	0.00170
4	1	8-1	0.1	0.95	logsig-purelin	0.00125
5	1	10-1	0.1	0.95	logsig-purelin	0.00099

In figure 5, it can be seen that there were 10 neurons an 1 *hidden layer* which have the smallest error (MSE) of 0.00099. The test was performed by using the best architectural design that has been obtained from the training, which was by using a network structure consisted of one layer consisting 6 *input* neurons. The first hidden layer consists of 10 neurons and the *output* layer consists of 1 neuron. The activation functions used were the binary sigmoid (*logsig*) and identity (*purelin*) functions. Result of data training in hybrid model in this paper is attached in Figure 4.

Fig 5. The result of data training phase on on Hybrid Arima-BPNN Model.

Figure 6 shows a regression of 0.98498 which means that there is a good correlation between the actual variables and the ARIMA-BPNN model developed in the test. The correlation size of 0.98498 indicates a good association between the related variables, they are; Monthly period, Selling Price, Sales History, Retail Sales Estimation, Project Sales Estimation, Number of Competitors.

Fig 6. The results of the test on Hybrid Arima-BPNN Model.

The regression results show the validity of performance of the generated ARIMA-BPNN network. The validation results show that the generated network output is almost equal to the target line and the data patterns are mostly close to the fit line, which indicate the data to be input into the network can be recognized by the network well.

3.3 Result of Forecasting

Table 3 below shows the comparison between actual data and hybrid model test. There are no significant differences seen from factual data differences and small target test data. The level of accuracy can be concluded into good category because it shows the data verification between the actual data and ANN which most of the data do not have significant differences.

Table 3. The comparasion between the original target and the test target.

Month	Target	Results of Hybrid Model Test
May 2016	50	69,53
June 2016	200	134,94
July 2016	100	55,93
August 2016	50	62,71
September 2016	50	51,61
October 2016	100	105,68
November 2016	50	54,57

4 Conclusion

The results obtained from the comparison of actual data with the hybrid model test is there is no significant differences seen from the difference of factual data and small target test data. The artificial neural network architecture that can be used for forecasting material inventory was the *multilayer feedforward* network with the best neuron generated was 6-10-1, or 6 input neurons, 10 with 1 (one) hidden layer. The input neuron consists of the Month of sales period, product price, sales history, estimated number of renovation projects, estimated number of new construction projects and number of competitors. The output neuron is the forecasting value of the inventory. The produced hybrid architectural model had 99% accuracy with an MSE of 0.00099926 on epoch 975 and training period of 00:00:01. For further development, this research can be conducted by collecting more data and longer time spans so that the scope of the research can be expanded.

The researcher would like to express his sincere gratitude to the Directorate of Research and Community Service of the Krida Wacana Christian University (Ukrida) for financial support from the acquisition of a research grant (under contract No. 01/UKKW/LPPM-FTIK/Lit/IV/2017) in 2017.

References

1. A.R.B. Albarune and M.M. Habib., **4** (2015)
2. B.K. Mishra, S. Raghunathan and X.H. Yue., *Production & Operations Management*, **18**, (2009)
3. A.Chetko J, V.Pandian M, S.Mile J, and D.georgi M, *World Automation Congress – Budapest Hungary*, (2006)
4. M. Khasel, M. Bijari, and G.A.R. Ardali, *Improvement of Auto- Regressive Integrated Moving Average models using Fuzzy logic and control* (2009)
5. G.P Zhang, *Neurocomputing*, **50** (2013)
6. L.Xiong and Y.Lu, *3rd International Conference on Information Management*, (2017)
7. R. Fildes, K. Nikolopoulos, S.F. Crone, A.A. Syntetos., *Journal of the Operational Research Society*, **50** (2008)
8. A. Ghosh and M. Chakraborty, *Int. J. Comput. Appl.*, **60** (2012)
9. W. Wu, W. Zhang, Y. Yang, and Q. Wang, in *Proc. 2nd Int. Conf. Softw. Eng. Data Mining (SEDM)*, **2** (2010)
10. A. Ozozen, G.Kayakutlu, M.Ketterer, and O.Kayalica, *Proceedings of PICMET '16 Technology Management for Social Innovation*, (2016)