Construction of ANFIS Model Based on LM-Test for Forecasting of Chili Price Data in Semarang

Tarno[1*], Di Asih I Maruddani[2], Rita Rahmawati[3]

{tarno.stat@gmail.com}

Department of Statistics, Universitas Diponegoro, Semarang 50275, Indonesia[1,2,3]

Abstract. The research aim is constructing Adaptive Neuro-Fuzzy Inference System (ANFIS) model for forecasting time series data. The ANFIS model is constructed and applied to chili price data in Semarang. The daily data are written during December 2018 to May 2019. The input selection in ANFIS is done by using the Lagrange Multiplier (LM) test. The lag-1 with 2 membership functions is selected as optimal input. The performance of prediction based on in-sample data is measured by the values of mean absolute percentage error (MAPE) and root mean squares error (RMSE). The values of MAPE and RMSE are 2.9% and 939.8 respectively.

Keywords: ANFIS, chili price data, forecasting, LM-test.

1   Introduction

Commodity price data include chili price data are usually non-stationary and non-linear which contain uncertainty characteristics [1-2]. Autoregressive integrated moving average (ARIMA) has been implemented for forecasting time series data [3]. The volatility models had been proposed by many researchers such as Engle (1982) and Bollerslev (1986). Autoregressive conditional heteroscedasticity (ARCH) for modelling inflation in UK developed by Engle [4]. The volatility model that called generalized autoregressive conditional heteroscedasticity (GARCH) proposed by Bollerslev [5]. Application of hybrid ARIMA-GARCH has been done in research for non-linear and non-stationary data [6]. Unfortunately, the model still has an advantages, because it cannot capture the uncertainty when implemented for non-linear and non-stationary data [7, 8].

The hybrid model between neural networks (NN) and fuzzy inference system (FIS) has been constructed for forecasting the non-linear time series model [9]. The combination model is called Adaptive Neuro-Fuzzy Inference System (ANFIS). There are many fields of time series research such as application of ANFIS based on singular spectrum analysis for forecasting chaotic time series [10]; chaotic time series prediction using improved ANFIS [11]; fuzzy time series forecasting [12]; developing a new approach for forecasting the trends of oil price [13]; forecasting of stock return [14]; forecasting of financial volatility [15]. The conclusion of the research related to ANFIS performance for forecasting non-linear data.

The crucial steps in ANFIS modeling are related to: (1) selecting inputs; (2) determining

the membership functions (number of clusters); and (3) generating rules. This research focused on the crucial procedure in ANFIS modelling which inspired by White. He proposed new procedure based on LM-test for selecting input in NN model [16]. The procedure was applied to construct ANFIS model of chili price data in Semarang.

This manuscript is organized as follows: materials of research which cover a structure of ANFIS was discussed in Section 2; Section 3 explains about the steps of modelling based on LM-test; Section 4 describes the results and discussion; and conclusion was presented in Section 5.

2   Materials

The materials and the sources are used in this study cover all articles which discuss about time series analysis using neural networks (NN), fuzzy inference system (FIS) and its hybrid. Based on literature review we can summarize about construction and application of ANFIS model for time series data. Before we discuss about procedure of ANFIS modeling, the most important material should be described in this section is the structure of ANFIS networks.

The NN architecture which applied in ANFIS consist of 5 fixed-layers [9, 17, 18]. Without loss of generality, the architecture of ANFIS for modeling time series data given 2 input variables Z_{t-1}, Z_{t-2} and single output variable Z_t by assuming rule-bases of Sugeno order-one with 2 rules is as follows:

If Z_{t-1} is A_{11} and Z_{t-2} is A_{21} then $Z_t^{(1)} = \theta_{11}Z_{t-1} + \theta_{12}Z_{t-2} + \theta_{10}$;

If Z_{t-1} is A_{12} and Z_{t-2} is A_{22} then $Z_t^{(2)} = \theta_{21}Z_{t-1} + \theta_{22}Z_{t-2} + \theta_{20}$;

where Z_{t-k} is A_{kj} as premise section, whereas $Z_t^{(j)} = \theta_{j0} + \sum_{k=1}^{2} \theta_{jk}Z_{t-k}$ as consequent section; θ_{jk}, θ_{j0} as linear parameters; A_{kj} as nonlinear parameter; $j = 1,2$; $k = 1,2$. If the firing strength for m values $Z_t^{(1)}, Z_t^{(2)}$ are w_1, w_2 respectively then the output Z_t can be expressed as:

$$Z_t = \frac{w_1 Z_t^{(1)} + w_2 Z_t^{(2)}}{w_1 + w_2}$$

$$Z_t = \bar{w}_1 Z_t^{(1)} + \bar{w}_2 Z_t^{(2)}.$$
$$(1)$$

The structure of ANFIS networks (Figure 1) have 5 layers and can be explained as follows [17]:

Layer 1: Each neuron in this layer is adaptive to the parameters of an activation function. The output of each neuron is the membership degree of input. For example the membership function of Generalized Bell is as follows:

$$\mu(Z_t) = \frac{1}{1+\left|\frac{Z_t - c_i}{a_i}\right|^{2b_i}}$$

where Z_t is input and a_i, b_i and c_i are premise parameters.

Layer 2: Each neuron in this layer is a permanent neuron that is given the symbol Π which is the product of all inputs in layer 1:

$$w_i = \mu_{Ai}(Z_{1t}) \times \mu_{Bi}(Z_{2t}), \ i = 1, 2.$$

Each neuron output is called the firing strength of a rule.

Layer 3: Each neuron in this layer is a fixed neuron with the symbol N which is the result of calculating the ratio of the i -firing strength to the total number of firing strengths in the second layer as follows: $\overline{w}_i = \frac{w_i}{w_1 + w_2}, i = 1, 2.$

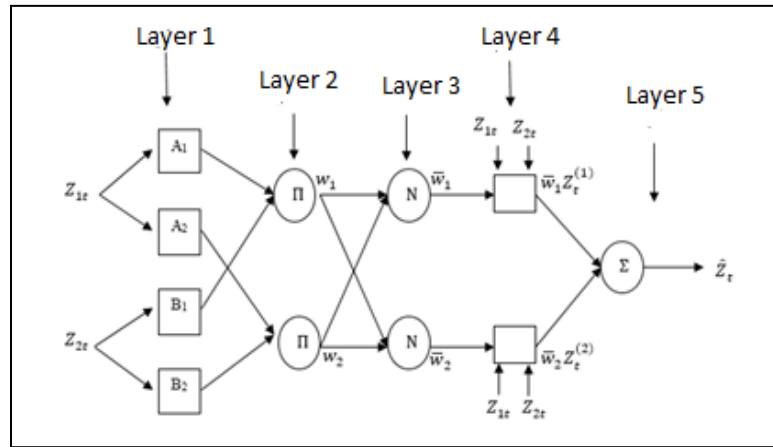The results of calculations at this layer are called normalized firing strength.

Figure 1. Structure of ANFIS Networks for Time Series Modeling

Layer 4: This layer is a neuron which is an adaptive neuron to an output:

$\overline{w}_i Z_t^{(i)} = \overline{w}_i (p_i Z_{1t} + q_i Z_{2t} + r_i)$

where $\overline{w}_i$ is normalized firing strength in the third layer while p_i, q_i, and r_i are parameters in these neurons called consequent parameters.

Layer 5: This layer is a single neuron with the symbol Σ which is the sum of all outputs from the fourth layer, as follows:

$$\hat{Z}_t = \sum_i \overline{w}_i Z_t^{(i)} = \frac{\sum_i w_i Z_t^{(i)}}{\sum_i w_i}$$

3 Methods

Lagrange Multiplier (LM) test is used for testing hypothesis of adding variables in ANFIS modeling. Some variables should be included to the model because the new inputs, number of clusters or number of rules to be added in the model.

3.1 Procedure for adding input variables

Determining input can be identified by a lag plot of data or the partial autocorrelation function (PACF) plot. The lag plot can also be used for testing linearity. The lag plot or PACF plot can be used for identification of autoregressive (AR) input. Based on the lag plot and PACF plot, the significant lags should be tested formally as input of ANFIS. Determining input is performed by constructing models which involve a number of input variables with minimum number of clusters and minimum number of rules. Firstly, constructing the models with 1 input, 2 clusters and 2 rules and then selecting the model which has the largest R^2. Adding input is evaluated based on LM test procedure to get a model with optimal inputs. If given 3 input variables $Z_{t-1}, Z_{t-2}, Z_{t-3}$ with 2 number of clusters, then the restricted model for this case can be written as [17, 18]:

$$Z_t = \theta_{11}(\bar{w}_1 Z_{t-1}) + \theta_{12}(\bar{w}_1 Z_{t-2}) + \theta_{10}\bar{w}_1 + \theta_{21}(\bar{w}_2 Z_{t-1}) + \theta_{22}(\bar{w}_2 Z_{t-2}) + \theta_{20}\bar{w}_2 + \varepsilon_t$$
(6)

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. While the unrestricted model for adding one input variable Z_{t-3} is as follow:

$$Z_t = \theta_{11}(\bar{w}_1 Z_{t-1}) + \theta_{12}(\bar{w}_1 Z_{t-2}) + \theta_{13}(\bar{w}_1 Z_{t-3}) + \theta_{10}\bar{w}_1 + \theta_{21}(\bar{w}_2 Z_{t-1}) + \theta_{22}(\bar{w}_2 Z_{t-2}) + \theta_{23}(\bar{w}_2 Z_{t-3}) + \theta_{20}\bar{w}_2 + \varepsilon_t$$
(7)

where $\vartheta_t \sim N(0, \sigma_\vartheta^2)$.

The formula of null hypothesis for adding variable Z_{t-3} is as follow:

$$H_0: \theta_{13} = \theta_{23} = 0$$

The steps of hypothesis test using LM-test are as bellow:

Step-1: Estimating the parameters of restricted model: $\hat{\theta}_{11}, \hat{\theta}_{12}, \hat{\theta}_{10}, \hat{\theta}_{21}, \hat{\theta}_{22}, \hat{\theta}_{20}$.

Step-2: Determining residuals

$$\hat{\varepsilon}_r = Z_t - \hat{\theta}_{11}(\bar{w}_1 Z_{t-1}) - \hat{\theta}_{12}(\bar{w}_1 Z_{t-2}) - \hat{\theta}_{10}\bar{w}_1 - \hat{\theta}_{21}(\bar{w}_2 Z_{t-1}) - \hat{\theta}_{22}(\bar{w}_2 Z_{t-2}) - \hat{\theta}_{20}\bar{w}_2.$$

Step-3: Fitting regression:

$$\hat{\varepsilon}_r = \theta_{11}(\bar{w}_1 Z_{t-1}) + \theta_{12}(\bar{w}_1 Z_{t-2}) + \theta_{13}(\bar{w}_1 Z_{t-3}) + \theta_{10}\bar{w}_1 + \theta_{21}(\bar{w}_2 Z_{t-1}) + \theta_{22}(\bar{w}_2 Z_{t-2}) + \theta_{23}(\bar{w}_2 Z_{t-3}) + \theta_{20}\bar{w}_2,$$ and determining the value of $LM = n \times R_{\hat{\varepsilon}}^2$.

The distribution of statistics LM is chi square with degree freedom 2.

Step-4: Making conclusion that adding variable Z_{t-3} to restricted model is rejected or received.

3.2 Procedure for adding number of clusters (membership functions) of inputs

The number of membership function of each input can be added by using LM-test when optimal inputs of ANFIS had been selected. If given 2 input variables Z_{t-1}, Z_{t-2} with 2 clusters, then the restricted model can be formulated as [17, 18]:

$$Z_t = \theta_{11}(\bar{w}_1 Z_{t-1}) + \theta_{12}(\bar{w}_1 Z_{t-2}) + \theta_{10}\bar{w}_1 + \theta_{21}(\bar{w}_2 Z_{t-1}) + \theta_{22}(\bar{w}_2 Z_{t-2}) + \theta_{20}\bar{w}_2 + \varepsilon_t$$
(8)

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ and unrestricted model for adding by one cluster is

$$Z_t = \theta_{11}(\bar{w}_1 Z_{t-1}) + \theta_{12}(\bar{w}_1 Z_{t-2}) + \theta_{10}\bar{w}_1 + \theta_{21}(\bar{w}_2 Z_{t-1}) + \theta_{22}(\bar{w}_2 Z_{t-2}) + \theta_{20}\bar{w}_2 + \theta_{31}(\bar{w}_3 Z_{t-1}) + \theta_{32}(\bar{w}_3 Z_{t-2}) + \theta_{30}\bar{w}_2 + \varepsilon_t$$
(9)

where $\vartheta_t \sim N(0, \sigma_\vartheta^2)$.

Formulation of testing hypothesis for adding one cluster of input variables can be written as:

$$H_0: \theta_{30} = \theta_{31} = \theta_{32} = 0$$

Procedure LM test of testing hypothesis are as the follows:

Step-1: Estimating the parameters of model (9): $\hat{\theta}_{11}, \hat{\theta}_{12}, \hat{\theta}_{10}, \hat{\theta}_{21}, \hat{\theta}_{22}, \hat{\theta}_{20}, \hat{\theta}_{31}, \hat{\theta}_{32}, \hat{\theta}_{30}$

Step-2: Determining the residuals:

$$\hat{\varepsilon}_r = Z_t - \hat{\theta}_{11}(\bar{w}_1 Z_{t-1}) - \hat{\theta}_{12}(\bar{w}_1 Z_{t-2}) - \hat{\theta}_{10}\bar{w}_1 - \hat{\theta}_{21}(\bar{w}_2 Z_{t-1}) - \hat{\theta}_{22}(\bar{w}_2 Z_{t-2}) - \hat{\theta}_{20}\bar{w}_2.$$

Step-3: Fitting regression:

$$\hat{\varepsilon}_r = \theta_{11}(\bar{w}_1 Z_{t-1}) + \theta_{12}(\bar{w}_1 Z_{t-2}) + \theta_{10}\bar{w}_1 + \theta_{21}(\bar{w}_2 Z_{t-1}) + \theta_{22}(\bar{w}_2 Z_{t-2}) + \theta_{20}\bar{w}_2 + \theta_{31}(\bar{w}_3 Z_{t-1}) + \theta_{32}(\bar{w}_3 Z_{t-2}) + \theta_{30}\bar{w}_3,$$ and determining the value of statistics $LM = n \times R_{\hat{\varepsilon}}^2$. Statistics LM has chi square distribution with degree freedom 3.

Step-4: Making conclusion that adding 1 cluster to inputs of restricted model is rejected or received.

3.3 Procedure for generating rule bases

The rule-bases of ANFIS are generated based on the number of optimal inputs and membership functions. There are m^p possible combination of rules to be generated when given m membership functions and p inputs. For instance, given 3 inputs $Z_{t-1}, Z_{t-2}, Z_{t-3}$ with 2 number of clusters, and one output Z_t by assuming first order Sugeno fuzzy system with 2 rules [17, 18]:

R1: If Z_{t-1} is A_{11} and Z_{t-2} is A_{21} and Z_{t-3} is A_{31} then $Z_t^{(1)} = \theta_{11}Z_{t-1} + \theta_{12}Z_{t-2} + \theta_{13}Z_{t-3} + \theta_{10}$;

R2: If Z_{t-1} is A_{12} and Z_{t-2} is A_{22} and Z_{t-3} is A_{32} then $Z_t^{(2)} = \theta_{21}Z_{t-1} + \theta_{22}Z_{t-2} + \theta_{23}Z_{t-3} + \theta_{20}$;

then the output of ANFIS is given by equation:

$$Z_t = \sum_{j=1}^{2}\left[\sum_{k=1}^{3}\theta_{jk}(\bar{w}_j Z_{t-k})\right] + \sum_{j=1}^{2}\theta_{j0}\bar{w}_j.$$

4 Results and Discussion

In this research, the monthly data of chili price in Semarang from December 2018 to May 2019 (www.hargajateng.org) are used for constructing the ANFIS model. Both time series and autocorrelation function (ACF) plots show that the price data are non-stationary (Figure 2a, 2b).
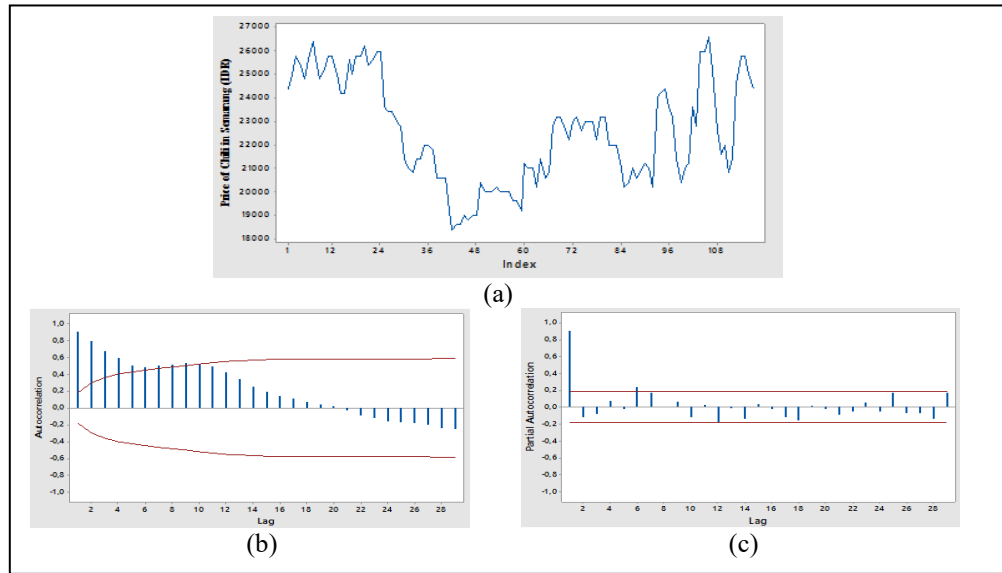
Figure 2(a). Time Series Plot; (b). Autocorrelation Function (ACF) Plot;
(c). Partial Autocorrelation Function (PACF) Plot

Based on partial autocorrelation function (PACF) plot (Figure 2c), the input ANFIS can not be decided. In this study, input variables in ANFIS to be selected by using LM test. Firstly, variables lag-1, lag-2, lag-3 with 2 clusters (membership functions) and its combination may be selected as inputs. The results of input selection are shown in Table 1.

Table 1. Result of LM test for input selection

Input	R_Z^2	R_e^2	LM	p-value
Z_{t-1}	0.8253	-	-	-
Z_{t-2}	0.6523	-	-	-
Z_{t-3}	0.4983	-	-	-
Z_{t-1}, Z_{t-2}	0.8260	0.0204	2.3295	0.127
Z_{t-1}, Z_{t-3}	0.8250	0.0217	2.4730	0.116

Based on Table 1, we conclude that lag-1 with 2 membership functions is considered as input ANFIS. The optimal input will be evaluated for determining the optimal number of membership

functions by using LM test. The result of membership functions selection is shown on Table 2.

Table 2. Result of LM test for membership functions selection

Input	MFs	R_e^2	LM	p-value
Z_{t-1}	3	0.0286	3.2573	0.071

According to the results on Table 2 variable input lag-1 with 3 membership functions cannot be selected as input of ANFIS because the p-value of the result test is greater than level of significance 0.05. So the variable lag-1 with 2 clusters is selected as input optimal of ANFIS. If given one input Z_{t-1} with 2 number of clusters, and one output Z_t by assuming first order Sugeno fuzzy system with 2 rules:

Rule-1: If Z_{t-1} is A_{11} then $Z_t^{(1)} = \theta_{11}Z_{t-1} + \theta_{10}$,
Rule-2: If Z_{t-1} is A_{12} then $Z_t^{(2)} = \theta_{21}Z_{t-1} + \theta_{20}$.
The result of optimal model is as follows:

$$Z_t = 0.8\bar{w}_{1,t}Z_{t-1} + 5656.8\bar{w}_{1,t} + 0.8\bar{w}_{2,t}Z_{t-1} + 4533.3\bar{w}_{2,t}$$
$$(10)$$

where $\bar{w}_{1,t} = \frac{w_{1,t}}{w_{1,t}+w_{2,t}}$, $\bar{w}_{2,t} = \frac{w_{2,t}}{w_{1,t}+w_{2,t}}$,

$$w_{1,t} = exp\left[-\frac{1}{2}\left(\frac{Z_{t-1}-24806}{2062.5}\right)^2\right],$$

$$w_{2,t} = exp\left[-\frac{1}{2}\left(\frac{Z_{t-1}-20772}{2284.4}\right)^2\right].$$

The MAPE and RMSE values of forecasting for data using Eq.10 are 2.9% and 939.8 respectively. The predicted value of chili price data based on model (10) can be seen in Figure 3.
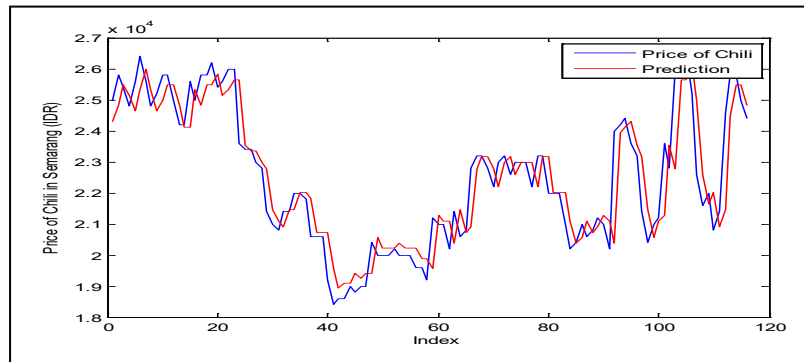

Figure 3. Chili Price Data and Its Prediction

5 Conclusion

Based on the result and discussion in the previous section, we concluded that the best ANFIS model has one input lag-1 with two membership functions and two rules. The predicted value of in sample data gave MAPE value 2.9% which is an excellent result.

References

[1]   Tsay, R.S.; 2005; *Analysis of Financial Time Series,* second Edition; Wiley Interscience, A John Wiley and Sons. Inc. Publication, USA.

[2]   Wei, W.W.S.; 2006; *Time Series Analysis: Univariate and Multivariate Methods*, Second Edition; Pearson Education Inc. Boston.

[3]   Box,G.E.P. , Jenkins, G.M. and Reinsel, G.C.; 1994; *Time Series Analysis, Forecasting and Control*, 3rd edition; Englewood Cliffs: Prentice Hall.

[4]   Engle, R.F.; 1982; Autoregressive Conditional Heteroscedasticity with Estimates of Variance United Kingdom Inflation, *Econometrica*, Vol. 50, No.4, pp.987-1007.

[5]   Bollerslev T.; 1986; Generalized Autoregressive Conditional Heteroscedasticity; *Journal of Econometrics*; 31, 307-327

[6]   Fahimifard, S.M., Homayounifar, M., Sabouhi, M. and Moghaddamnia, A.R.; 2009; Comparison of ANFIS, ANN, GARCH and ARIMA Techniques to Exchange Rate Forecasting; *Journal of Applied Sceinces* 9, pp.3641-3651.

[7]   Talebizadeh, M. and Moridnejad, A.; 2011; Uncertainty analysis for the forecast of lake level fluctuations using ensembles of ANN and ANFIS models; *Expert Systems with Applications*, 38, 4126-4135

[8]   Haykin; 1999; *Neural Networks: A Comprehensive Foundation*, Second Edition; Prentice Hall International, Inc..

[9]   Jang J-S R, Sun C-T, Mizutani E.; 1997; *Neuro-Fuzzy and Soft Computing: A computational Approach to Learning and Machine Intelligent*; Prentice Hall International.

[10]   Daneshmand H, Tavousi T, Khosravi M, Tavakoli S; 2015; Modeling minimum temperature using adaptive neuro-fuzzy inference system based on spectral analysis of climate indices: A case study in Iran; *Journal of the Saudi Society of Agricultural Sciences*, 14 , pp.33-40

[11]   Abdollahzade M, Miranian A, Hossein H, Iranmanesh H.; 2015; A new hybrid enhanced local linear neuro-fuzzy model based on the optimized singular spectrum analysis and its application for nonlinear and chaotic time series forecasting; *Information Sciences*, 295, pp. 107-125

[12]   Behmanesh M, Mohammadi M, Naeini V S; 2014; Chaotic Time Series Prediction using Improved ANFIS with Imperialist Competitive Learning Algorithm; *International Journal of Soft Computing and Engineering (IJSCE)*, Volume 4- Issue 4 ,pp. 25-33

[13]   Cheng, S.-H., Chen, S.-M., Jian, W.-S.; 2016; Fuzzy time series forecasting based on fuzzy logical relationships and similarity measures, *Information Sciences*, 327, pp. 272-287

[14]   Cheng, C. H. and Wei, L. Y.; 2010; One step-ahead ANFIS time series model for forecasting electricity loads; *Optimization and Engineering*, 11, 303-317.

[15]   Luna, I. and Ballini, R.; 2012; Adaptive fuzzy system to forecast financial time series volatility;

Journal of Fuzzy and Intelligent Systems 23, pp. 27-38.

[16]   Anders, U and Korn, O.; 1999; Model Selection in Neural Networks; *Neural Networks*, 12, 309-323

[17]   Tarno T, Suparti S, Ispriyanti D.; 2018; Modelling of Cayenne Production Data in Central Java Using ANFIS, *Model Assisted Statistics and Applications* 13(1) p 45-52

[18]   Tarno, Rusgiyono, A., and Sugito; 2019; Adaptive Neuro Fuzzy Inference System (ANFIS) approach for modeling paddy production data in Central Java, *Journal of Physics: Conference Series, Volume 1217, conference 1* p 1-8