Adaptive Neuro-Fuzzy Inference System implementation for farmer’s term of trade forecasting in West Sumatra

Y Aufar¹ and I S Sitanggang¹

¹ Department of Computer Science, Faculty of Natural Science and Mathematics, Bogor Agricultural University (IPB), Jl. Meranti Wing 20 Level V, Kampus IPB Darmaga, Bogor, Indonesia 16680, e-mail: imas.sitanggang@ipb.ac.id

Abstract. The farmer’s term of trade (NTP) is used by the Central Bureau of Statistics (BPS) as one of the indicators for measuring the level of welfare or purchasing power of farmers. To prepare preventive measures when the NTP index falls from the previous period, the relevant parties need to predict NTP for the coming period. The purpose of this study is to measure the performance of Adaptive Neuro-Fuzzy Inference System (ANFIS) in predicting NTP of West Sumatra Province in the coming month. The data used are monthly NTP data of West Sumatra Province obtained from the BPS website www.sumbar.bps.go.id from 2013 to 2016 for the network training process. Model evaluation was done by comparing the test results with actual data in 2017. Forecasting systems are divided into two types, namely time series model and the multivariate model. The test results show that the time series model has the smallest RMSE of 0.3430 for the training set and the RMSE value is 1.4570 for the testing set.

Keywords: ANFIS, farmer’s term of trade, forecasting.

1. Introduction
Agriculture is still one of important sectors in the economy of West Sumatra, where each year has a percentage of more than 20 percent in the Gross Regional Domestic Product (PDRB) of West Sumatra Province [1]. This sector is expected to increase farmers’ income and can help alleviate poverty. Therefore, the role of the government and related parties in carrying out strategic policies to improve the welfare of farmers is required. The Central Statistics Agency (BPS) has published data on the farmer’s term of trade (NTP) as an indicator that can measure the level of welfare of farmers. NTP describes the ratio between the price index received by farmers (It) and the price index paid by farmers (Ib) in percentage. The higher the NTP, the more prosperous the level of life of the farmer. The right method is needed to find out or predict how many NTP indices will be in the coming period so that action can be taken if the NTP index has decreased from the previous period.

Some methods that are often used in time series data forecasting are neural networks [2] and fuzzy logic [3]. Neural networks have the advantage of computing by simulating structures and functions such as performance in brain nerve tissue. The advantages of the neural network can be used to deal with conditions with increasing uncertainty [4]. Whereas in fuzzy logic, it has advantages in dealing with various uncertain or unclear conditions [5]. Besides having advantages, of course, neural networks and fuzzy logic also have disadvantages. Neural networks have deficiencies in complex processing stages so that they are not effective on networks which are large enough and do not have reasoning capabilities.
like those in fuzzy logic. Whereas in fuzzy logic, it has the disadvantage of not having the ability to learn and adapt like a neural network. Therefore, the Adaptive Neuro-Fuzzy Inference System (ANFIS) method is developed which combines the two methods [6].

In this study, the ANFIS method is used to predict NTP in West Sumatra Province. The use of the ANFIS method in the prediction process is expected to produce a more accurate predictive value. ANFIS analysis in this study uses the order-1 Sugeno model. Order-1 is chosen with consideration of simplicity and ease of calculation. The learning algorithm used is a hybrid optimization method. The software used is MATLAB.

2. Related research
NTP-related research in Indonesia was made by Hendayana using regression to determine the relationship between factors influencing the NTP in North Sumatra. Hendayana said that NTP is a relationship between agricultural products sold with other goods and services purchased by farmers [7]. Another research on NTP for time series datasets was conducted by Khusniyah and Sutikno; they studied the predictions of NTP in East Java Province using backpropagation neural network algorithms. The smallest percentage error or the highest accuracy is obtained when the number of hidden layer nodes is 7, and the learning rate value is 0.1. But neural networks are not effective on large enough networks because they require long and complicated processing times [8].

Related research on ANFIS method has been widely applied to time series data prediction problems. Putra et al. applying the ANFIS method to predict rainfall in Kalimantan [9]. Faulina et al. applied the ANFIS method and compare it with the Autoregressive Integrated Moving Average (ARIMA) method for predicting monthly rainfall data at certain area in East Java Province, Indonesia, namely Pujon and Wagir area. Based on the Root Mean Squared Error (RMSE) in the test dataset, the results showed that the ANFIS method produced an estimate of the monthly data of rainfall in the Pujon area more accurately while the ARIMA model produced a good estimate of monthly rainfall data in the Wagir area [10]. Wei applied the ANFIS method in the case of TAIEX stock forecasting [11]. Dewi and Himawati used the ANFIS method to predict the unemployment rate [12]. Forecasting the level of atmospheric turbidity by the ANFIS method was carried out by Nou et al. [13]. The studies carried out show that the ANFIS method approach is well known and widely used as a time series data prediction algorithm. Therefore, this study uses ANFIS as a method to predict NTP in the Province of West Sumatra with the results of the right and accurate NTP prediction.

3. Methodology
There are several stages of the process to predict NTP in West Sumatra Province by using ANFIS method. The process stages are illustrated in figure 1.
Figure 1. Research Methodology.

3.1. Data Collection and Processing
The data used in this study are secondary data obtained indirectly through intermediate media. This study uses NTP data in West Sumatra from 2013 to 2017 obtained from the website of the Central Bureau of Statistics (BPS) www.sumbar.bps.go.id and accessed on April 14, 2018. The data obtained are divided into two parts, first as network training data, and the second as test data. Training data contain the West Sumatra NTP data in the period of 2013, 2014, 2015, and 2016. Testing data contain the West Sumatra NTP data in 2017.

3.2. Formation of ANFIS network
In this study, in general, the monthly NTP forecasting system is made to predict future NTP. Based on the data, this NTP forecasting system uses the ANFIS method with two input and output models, namely the time series model where all input and output variables are the same variables and multivariate models with different input variables with output variables [14]. ANFIS time series model, for example, to predict NTP, the input used is current NTP and NTP data in one month ago, while the output is NTP data in one month ahead. The ANFIS model is a double variable, the type of input data variable is different from the type of output variable, for example, to predict the NTP in one month to come, the input used is the current data of the Farmer Accepted Price Index (It) and the Price Index Paid by Farmers
There are several input-output combinations that are formed to get the best predictor of performance. The following is the input-output arrangement that was made as an NTP predictor, as given in Table 1 and Table 2.

| Input 1                           | Input 2                           | Output                           |
|----------------------------------|----------------------------------|----------------------------------|
| Farmer Exchange Rate this month (t) | Farmer Exchange Rate one month ago (t-1) | Farmer Exchange Rate next month (t + 1) |

| Input 1                           | Input 2                           | Output                           |
|----------------------------------|----------------------------------|----------------------------------|
| Price Index Received by Farmers this month (t) | Price Index Paid by Farmers this month (t) | Farmer Exchange Rate next month (t + 1) |

4. Determination of Membership Function and Number of Clusters
A neuro-fuzzy program implementation uses the ANFIS method. ANFIS trains FIS with FIS initialization, which sets the initial price of the membership function parameters in the FIS. FIS initiation includes selection of membership functions, selection of membership function types, and selection of training iterations (epoch).

4.1. ANFIS Training
In the program development phase, training was carried out with ANFIS to find parameters in the IF section, namely parameters \( a \) and \( c \) (premise parameters), and parameters in the THEN section (consequent parameters). After these parameters are known with a small error, these parameters are used in the next stage. At this stage, we use Fuzzy C-Means (FCM) to do clustering to determine data assigned to a cluster. The next step is calculating the mean and standard deviation for each cluster, the mean value, and this standard deviation will then be the value of \( a \) and \( c \) of the first parameter value. Furthermore, the output in the first layer is the degree of membership of each data, in the second layer is the degree of membership multiplication, in the third layer is normalized, the fourth layer looks for consequent parameters, in the fifth layer is the sum of the fourth layer. Recursive least-squares estimation is used to correct consequent parameters until we get a small error value in the forward step, while in the reverse step we use error propagation to fix the premise parameters.

4.1.1. Fuzzyfication layer. After getting the results of fuzzy clustering we then calculate the first layer by calculating the mean and standard deviation. The Generalized Bell membership function is done with the following formula:

\[
\mu = \frac{1}{1 + \left| \frac{A - c}{a} \right|^2 b}
\]

where \( A \) as an input, \( A \{A_{1,t}, A_{2,t}\} \) and \( \{a, b, and c\} \) are parameters, usually \( b = 1 \). If the values of these parameters change, then the shape of the curve will change too. These parameters are usually called the premise parameters [15].

4.1.2. Second Layer. This layer is the result of the degree of membership from the first layer, as follows:

\[
w_t = \mu_{A_t} \cdot \mu_{B_t}
\]
The results of this calculation are called firing strength of a rule. Each neuron represents the rule to-
note $i$ [16].

4.1.3. Third Layer. Each neuron in this layer is a fixed neuron (given the symbol N) which is the
result of calculating the ratio of $i$ ($w_i$) firing strength to the total number of firing strengths in the
second layer, as follows:

$$\bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2$$

(3)

This result is also called normalized firing strength [16].

4.1.4. Fourth Layer. This layer is a neuron which is an adaptive neuron to an output, as follows:

$$\bar{w}_i \cdot f_i = \bar{w}_i (p_i Z_{1,t} + q_i Z_{2,t} + r_i)$$

(4)

$\bar{w}_i$ is normalized firing strength in the third layer and $p_i$, $q_i$, and $r_i$ are the parameters on the
neuron. These parameters are usually called consequent parameters [16].

4.1.5. Fifth Layer. This layer is a single neuron (symbol $\sum$) which is the sum of all outputs from the
fourth layer, as follows [17]:

$$\sum \bar{w}_i \cdot f_i = \frac{\sum \bar{w}_i \cdot f_i}{\sum \bar{w}_i}$$

(5)

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{ANFIS.png}
\caption{ANFIS Structure [18].}
\end{figure}

4.1.6. Recursive LSE. If the $m$ element is owned by the vector $Z_t$ ($Z_t$ measuring $m \times 1$) and $n$
parameters $\theta$ ($\theta$ measuring $n \times 1$), with the row $i$, in the matrix $[A; Z_t]$ is denoted as $[a_t^i ; Z_t]$. Least-
squares estimator is written as follows:

$$A^T A \hat{\theta} = A^T Z_t$$

(6)

If $A^T A$ is nonsingular and $\hat{\theta}$ unique, it can be given:

$$\hat{\theta} = (A^T A)^{-1} A^T Z_t$$

(7)

or by removing $\hat{\theta}$ and assumed the number of lines of $A$ and $Z_t$ pair is $k$ then it is obtained:

$$\theta_k = (A^T A)^{-1} A^T Z_t$$

(8)

there is a recursive LSE added a data pair $[a_t^i ; Z_t]$, so there are as many as $m+1$ pairs of data. Then LSE
$\theta_{k+1}$ is calculated with the help of $\theta_k$. Because there are as many parameters as $n$ then with the inversion
method, as follows:
\[ P_n = (A_n^T A_n)^{-1} \]  
and  
\[ \theta_n = p_n A_n^T Z_{t(n)} \]

Then the iteration starts from data \((n+1)\), with \(P_0 \) and \(\theta_0 \) is calculated by the equation \(P_n \) and \(\theta_n \), the value of \(P_{k+1} \) and \(\theta_{k+1} \) can be calculated as follows:

\[ P_{k+1} = P_k - \left( \frac{\alpha_k \alpha_{k+1} P_k}{1 + \alpha_k + \alpha_{k+1}} \right) \]

\[ \theta_{k+1} = \theta_k + \left( \frac{\alpha_k \alpha_{k+1} Z_{t(k+1)}}{1 + \alpha_k + \alpha_{k+1}} \right) \]

ANFIS trains FIS by modifying membership function parameters until a minimum error between FIS output and output training data is obtained [17].

4.2. Comparative Data with Prediction Results

The model validation is the FIS testing process that has been trained by ANFIS but uses input/output data that has not been trained by FIS.

5. Results and Discussion

5.1. ANFIS analysis on monthly time series data

5.1.1. Cluster Selection. By using the `gbellmf` (Generalized Bell) membership function, which is available in MATLAB, in each ANFIS training with a different number of clusters, a summary of the results is given in Table 3.

| ANFIS Input          | Number of Clusters | RMSE       |
|----------------------|--------------------|------------|
|                      | Training Data (in-sample) | Data Testing (out-sample) |
| NTP (t) and NTP (t-1) | [2 2]               | 0.6856     | 0.8605     |
|                      | [3 3]               | 0.6399     | 1.1279     |
|                      | [4 4]               | 0.5018     | 1.3868     |
|                      | [5 5]               | 0.3875     | 1.4022     |
|                      | [6 6]               | 0.3418     | 1.4768     |
|                      | [7 7]               | **0.3039** | **1.5436** |
|                      | [8 8]               | 0.1943     | 3.8771     |

According to Table 3, the best number of clusters is determined, namely those that produce RMSE on the smallest training and testing data. The results of testing the number of clusters show that up to the cluster [8 8], the RMSE value has decreased in training data. But there was a high increase in testing data. This happens because the data with the number of clusters [7 7] has exceeded the optimal state. Thus, in this experiment that the best input is NTP (t) and NTP (t-1) with the number of clusters [7 7].

5.1.2. Selection of Membership Functions. Based on the previous analysis, the best input is NTP (t) and NTP (t-1) with the number of clusters [7 7]. Then ANFIS training was carried out on NTP (t) and NTP (t-1) input with the number of clusters [7 7] on several different membership functions. The following results are obtained, as shown in Table 4.
Table 4. ANFIS training on monthly time series data based on membership functions

| Membership Function | RMSE |  
|---------------------|------|
| trimf | 0.4468 | 1.5177 |
| trapmf | 0.5963 | 1.5055 |
| gbellmf | 0.3039 | 1.5436 |
| gaussmf | 0.3430 | 1.4570 |

According to Table 4 the best membership function is determined, which produces the smallest RMSE on training and testing data. Although gbellmf (Generalized Bell) gives the smaller RMSE on training data than RMSE from the gaussmf (Gaussian), but for testing data, Gaussian is much better than Generalized Bell in resulting small RMSE. It is concluded that the best membership function is Gaussian.

5.1.3. Results Analysis. From the training conducted on the data, it is concluded that the best ANFIS models are NTP (t) and NTP (t-1) inputs with number of clusters [7 7] and the Gaussian membership function. RMSE on training data is 0.3430 and RMSE on testing data is 1.4570.

5.2. ANFIS analysis on multi-variable monthly data

5.2.1. Cluster Selection. By using the gbellmf (Generalized Bell) membership function, which is available in MATLAB, in each ANFIS training with a different number of clusters, a summary of the results is provided in Table 5.

Table 5. ANFIS training on monthly multi-variable data based on the number of clusters

| ANFIS Input       | Number of Clusters | RMSE |  
|-------------------|--------------------|------|
| It (t) and Ibf (t) | [2 2]              | 0.9238 | 35.8984 |
|                   | [3 3]              | 0.6320 | 34.2002 |
|                   | [4 4]              | 0.5864 | 32.9146 |
|                   | [5 5]              | 0.4961 | 31.6605 |
|                   | [6 6]              | 0.4865 | 28.5457 |
|                   | [7 7]              | 0.4798 | 32.0965 |
|                   | [8 8]              | 0.4241 | 27.0349 |

According to Table 5, the best number of clusters is determined, namely those that produce the smallest RMSE in training and testing data. It is concluded that the best inputs are It (t) and Ibf (t), with the best number of clusters [8 8].

5.2.2. Selection of Membership Functions. Based on the previous analysis, the best inputs are It (t) and Ibf (t) with the number of clusters [8 8]. Then the input of ANFIS training are It (t) and Ibf (t) with the number of clusters [8 8] on several different membership functions. The following results are obtained, as given in Table 6.
Table 6. ANFIS training on multi-variable monthly data based on function membership

| Membership Function | RMSE       |       |
|---------------------|------------|-------|
|                     | Training Data (in-sample) | Data Testing (out-sample) |
| trimf               | 0.5180     | 49.8184 |
| trapmf              | 0.6200     | 39.7706 |
| gbellmf             | **0.4241** | **27.0349** |
| gaussmf             | 0.4874     | 26.3147 |

According to Table 6, the best membership function is determined, which produces the smallest RMSE on training and testing data. Although the RMSE from Gaussian is smaller for data testing than the RMSE from gbellmf (Generalized Bell), but for Generalized Bell, RMSE on training data is much smaller than those of Gaussian. It is concluded that the best membership function is Generalized Bell.

5.2.3. Results Analysis. From the training conducted on the data, it is concluded that the best ANFIS model has the input It (t) and Ib (t) with the number of clusters [8 8] and the Generalized Bell membership function. RMSE on training data is 0.4241 and RMSE on testing data is 27.0349.

6. Conclusions and Suggestions

This study has successfully applied the fuzzy neuro model, namely the ANFIS on NTP data from January 2013-December 2017. Based on experimental results we conclude that the best model for inflation prediction data according to the smallest RMSE in data training and data testing is obtained on the time series monthly data with input NTP (t) and NTP (t-1), number of clusters [7 7], and the Gaussian membership function. On training data prediction of NTP using the model the ANFIS has RMSE of 0.3430 whereas on the testing data has RMSE of 1.4570. These results are still better than prediction of NTP for multi-variable monthly data with input It (t) and Ib (t), the number of clusters [8 8] and the Generalized Bell membership function which produce RMSE of 0.4241 on the training data and RMSE of 27.0349 on the testing data. RMSE in the testing data is very large, it can be concluded that the prediction of NTP using the ANFIS model is not suitable for multi-variable monthly data, but it is suitable for monthly time series data.

Based on the results of this study a suggestion can be considered in the further study. Further work is expected to be able to use other methods so that they can be compared in term of their performance.

References
[1] V. Wardi 1943 Nilai Tukar Petani Provinsi Sumatera Barat 2013-2017 T. Sugiyarto, Ichsan, Y. Buchari, M. H. Bahar, and Deswaty, Eds. Badan Pusat Statistik Provinsi Sumatera Barat
[2] W. S. McCulloch and W. Pitts 1943 A logical calculus of ideas immanent in nervous activity Bulletin of Mathematical Biophysics
[3] L. Zadeh 1965 Fuzzy sets Information and Control 8 pp. 338–353
[4] S. Hardwinarto, M. Aipassa et al. 2015 Rainfall monthly prediction based on artificial neural network: A case study in tenggarong station, east kalimantan-indonesia Procedia Computer Science 59 pp. 142–151
[5] M. Hasan, X. F. Shi, T. Tsegaye, N. U. Ahmed, and S. M. M. Khan 2013 Rainfall prediction model improvement by fuzzy set theory J. Water Resource and Protection 5 no. 01 p. 1
[6] J.-S. Jang 1993 Anfis: adaptive-network-based fuzzy inference system IEEE transactions on systems, man, and cybernetics 23 no. 3 pp. 665–685
[7] R. Hendayana 2001 Analisis faktor-faktor yang mempengaruhi nilai tukar petani Pusat Penelitian dan Pengembangan Sosial Ekonomi Pertanian, Badan Penelitian dan Pengembangan Pertanian Bogor
[8] T. W. Khusniyah and S. Sutikno 2016 Prediksi nilai tukar petani menggunakan jaringan syaraf tiruan backpropagation Scientific Journal of Informatics 3 no. 1 pp. 11–18
[9] B. S. Putra, R. S. Wahono et al. 2011 Simulasi penerapan anfis pada sistem lampu lalu lintas enam ruas Kursor 6 no. 2
[10] R. Faulina, D. A. Lusia, B. W. Otok, H. Kuswanto et al. 2012 Ensemble method based on anfis-arima for rainfall prediction Statistics in Science, Business, and Engineering (ICSSBE) International Conference on. IEEE pp. 1–4.
[11] L.-Y. Wei 2011 An expanded adaptive neuro-fuzzy inference system (anfis) model based on ar and causality of multi-nation stock market volatility for taiex forecasting 5 no. 15, p. 6377
[12] C. Dewi and W. W. Himawati 2015 Prediksi tingkat pengangguran menggunakan adaptif neuro fuzzy inference system (anfis) Proceedings Konferensi Nasional Sistem dan Informatika (KNS&I)
[13] J. Nou, R. Chauvin, A. Traoré, S. Thil, and S. Grieu 2014 Atmospheric turbidity forecasting using side-by-side anfis Energy Procedia 49 pp. 2387–2397
[14] N. Chasanah 2015 Perancangan sistem prediktor cuaca dengan metode anfis untuk menentukan produktivitas panen sayuran kubis putih (brassica oleracea var. capitata) di karangploso kabupaten malang Ph.D. dissertation Institut Teknologi Sepuluh Nopember,
[15] Tjahjono, E. Martiana, and T. H. Ardhinata 2011 Penerapan adaptive neuro fuzzy inference system (anfis) untuk sistem pengambilan keputusan distribusi obat pada sistem informasi terintegrasi puskesmas dan dinas kesehatan Electronic Engineering Polytechnic Institute of Surabaya (EEPIS) Indonesia 4 no. 1, pp. 338–344
[16] S. Sri Kusumadewi 2010 Neuro-fuzzy: Integrasi sistem fuzzy & jaringan syaraf Yogyakarta: Graha Ilmu
[17] S. Kusuma 2003 Artificial intelligence (teknik dan aplikasinya) edisi pertama Yogyakarta: Graha Ilmu
[18] J.-S. R. Jang, C.-T. Sun, and E. Mizutani 1997 Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [book review] IEEE Transactions on automatic control 42 no. 10 pp. 1482–1484