

# Interval Regression With Neuro-Fuzzy and Madaline Architecture for Prediction of Rice Production

Fatma Agus Setyaningsih

Department of Computer System  
 Faculty of Mathematics and Natural Sciences  
 Tanjungpura University  
 Pontianak, Indonesia  
 fatmasetyaningsih@siskom.untan.ac.id

**Abstract**— The paddy production interval value in Kubu Raya Regency can be predicted by using past values in conjunction with one of the neuro-fuzzy's methods, which is the interval regression model. Interval regression is a backpropagation-based method. There are two backpropagated-networks in this model, with one set to find a lower bound and the other set to find an upper bound. The constructed system has the capability to receive input data in the form of statistics, which in this case is the rice production data of the last three periods, and then processing them using the interval regression model via the neuro-fuzzy. This Madelin Method uses three layers, with four input nodes, three hidden nodes, and one output node, which then delivers the result in the form of a predictive value of future rice production values. Our primary goal is to predict next year's rice production. In this article, I used two machine prediction-based models, which is the interval regression model with neuro-fuzzy and the madaline. This research used 3 time periods to determine the best parameters for accuracy. To evaluate the performance of this system, we used the Mean Squared Error (MSE) measurement. The results gained indicated that the performance of the interval regression model with neuro-fuzzy worked better than the madaline with a learning parameter value of 0,09. The benefits of our method is that the regression interval method with neuro-fuzzy is one of the strongest machine prediction model in the world.

**Keywords**-interval regression; madaline architecture; neuro-fuzzy; rice production

## I. INTRODUCTION

Crops in the form of edible, seed-bearing grasses are widespread in Indonesia. Depending on how it's cultivated, paddy is divided into two types: dry paddy (*gogo*) which are planted in normal soil and field paddy which are planted in fields which are always kept wet [1].

While an essential food staple of Indonesian society, rice can't be produced every day. There are four factors which are critical to rice production: the harvesting area's size, the planting area's size, rate of precipitation and productivity.

The Kubu Raya Regency is an area with great potential for food self-sufficiency than other regions, due to support from farmers, agencies, and other elements of society. The keys to the Regency's success is due to its capacity for food security and self-sufficiency, maintained by using long and short-term planning. According to data from Kubu Raya Regency's *Badan*

*Pusat Statistik (BPS)*, Kubu Raya Regency produced a net rice production of 117,422 tonnes from a harvesting area of 59,945 hectares in 2014. Meanwhile, the current population of Kubu Raya Regency is estimated at around 537,207 people, with a 139.15 kg food consumption per capita a year. But Kubu Raya Regency's rice production at 2014 decreased by 5.79% due to bad weather conditions and shrinking harvesting area.

The regression interval method is used to assist people in making decision. Alternatively, this method is also called regression interval method with neuro-fuzzy. The regression interval method with neuro-fuzzy can generate upper limit intervals and lower limit intervals. [2]

The Madaline Neural Network is a multi-layered neural network algorithm, which means that it uses hidden layers with an activation function to optimize input values before it goes to the output layer. Forming a new Madaline network requires the combination of several Adaline networks. [3]

Prediction of harvesting area and rice production in the Banyumas Regency using the ANFIS method in 2011 and 2012 by using network architecture amounted to 236,619 hectares and 314,913 tonnes of rice in 2011, and 64,799 hectares and 326,839 tonnes of rice in 2012. The prediction was 31.22% accurate, according to MAPE criterias. [4]

Weather also plays a major role in rice production. If say the rainfall in a certain area was very low, it could cause paddies to not mature properly, resulting in failed harvests and consequently, decreased rice production. This would in turn raise the local price of rice, and increase the demand for rice imports. In order to alleviate this problem, it is vital that Kubu Raya Regency possess the ability to predict future rice production.

## II. RESEARCH METHOD

The research methodology describes the sequence of steps taken in the research and the methods used to solve the problem. An illustration of said steps can be seen in Figure 1.

### A. Problem Identification

At this stage, the problem is identified, which in this case, is how to predict the production of rice in Kubu Raya Regency at certain time periods by using the regression interval method with neural-fuzzy and madaline method. The time periods are

from January to April, May to August and September to December. The two aforementioned methods will then be compared by using the Mean Squared Error (MSE) parameters.

*1) Datasets*

The scope of this study focuses primarily on the Kubu Raya Regency area. Collecting data that influences rice production is done by non-formal interviews, or “face-to-face” Q&A sessions. The result is rice production data from 2009 to 2016, obtained from the BPS’s Kubu Raya office. The data inputted into the system are the four factors which affected rice production from 2009 to 2016. The factors are as follows:

- a) *Size of harvesting area.*
- b) *Size of plantation area.*
- c) *Productivity.*
- d) *Rainfall.*

The data that will be used is data from three time periods, which are:

- a) *January – April.*
- b) *May – August.*
- c) *September – December.*

105 data of various types were used in this research, such as the size of the planting area, the size of the harvesting area, productivity, rainfall, and the target value of rice production in Kubu Raya Regency.

This System will conduct separate training for each three time periods, with the rice production data in Kubu Raya Regency as the goal of this research. Said data will be split into two parts, that is rice production data from 2010 to 2014 as output data for training and data from 2015 to 2015 as output data for testing. The training done in this research is used to verify that system produces the desirable results, and can be used to perform prediction on the coming years for each of the aforementioned three time periods. The end result of this research is the accuracy of predicting the value of rice production for each three time periods in Kubu Raya Regency.

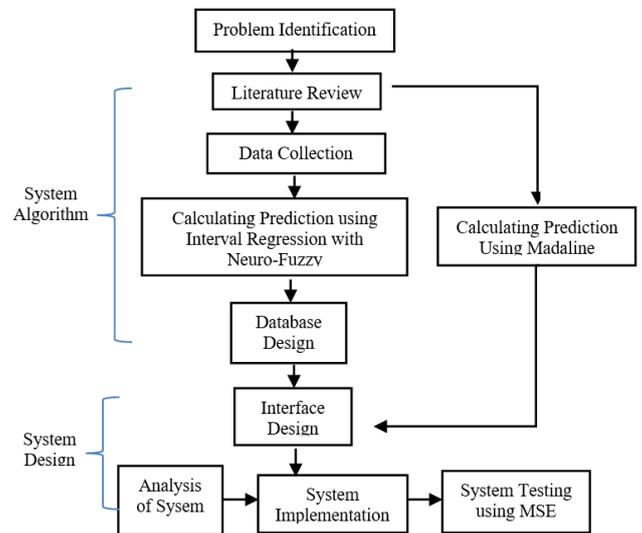


Fig. 1. Stages of Research

TABLE I. THE DATA USED TO PREDICT THE VALUE OF RICE PRODUCTION IN KUBU RAYA REGENCY

Year	Input Data	Period			Rice Production Target (ha)		
		January-April	May-August	September-December	1 <sup>st</sup> Time Period	2 <sup>nd</sup> Time Period	3 <sup>rd</sup> Time Period
2009	Size of Planting Area (ha)	145.152	14.157	40.053	128.469	18.389	41.311
	Size of Harvesting Area(ha)	44.338	3.883	11.382			
	Productivity (ku/ha)	32,74	36	35,19			
	Rainfall(mm)	2.552,5	2.515	6.451			
2010	Size of Planting Area (ha)	867	13.965	42.814	132.840	20.598	53.454
	Harvesting Area (ha)	38.833	5.128	11.835			
	Productivity (ku/ha)	33,08	35,86	34,91			
	Rainfall(mm)	4.366,5	7.444	6937			
2011	Size of Planting Area (ha)	3.293	17.737	40.415	134.531	23.610	46.860
	Size of Harvesting Area(ha)	39.460	5.749	15.485			
	Productivity (ku/ha)	33,66	35,83	34,34			
	Rainfall(mm)	2.613,5	3.584	5.378,5			
2012	Size of Planting Area (ha)	3.074	11.531	46.424	147.946	35.675	34.291
	Harvesting Area (ha)	39.709	6.677	13.646			
	Productivity (ku/ha)	33,88	35,36	34,34			

Year	Input Data	Period			Rice Production Target (ha)		
		January-April	May-August	September-December	1 <sup>st</sup> Time Period	2 <sup>nd</sup> Time Period	3 <sup>rd</sup> Time Period
	Rainfall(mm)	3.378,5	2.614,5	5.378,5			
2013	Size of Planting Area (ha)	8.955	13.469	50.295	170.502	14.344	20.441
	Harvesting Area (ha)	43.443	10.158	10.009			
	Productivity (ku/ha)	34,06	35,12	34,26			
	Rainfall(mm)	5.084	4.101	4721			
2014	Size of Planting Area (ha)	2.796	6.056	55.281	152.775	15.963	11.690
	Harvesting Area (ha)	49.871	4.097	5.977			
	Productivity (ku/ha)	34,19	35,01	34,2			
	Rainfall(mm)	5273	5.142	5292			
2015	Size of Planting Area (ha)	1.157	6.310	36226	152.394	14.496	27.885
	Harvesting Area (ha)	44.640	4.624	3.415			
	Productivity (ku/ha)	34.22	34.52	34.23			
	Rainfall(mm)	5056	5.039	5058,5			

*a) Harvesting Area Data*

Harvesting Area Data is used as input data in this research. The data used is data from three time periods: January to April, May to August and September to December, from 2009 to 2015. This data was obtained from Kalimantan Province’s Agriculture Department’s Agency of Crops and Horticulture as well as the Agency of Farming and Agriculture. This data is given the variable initial X1.

*b) Paddy Planting Area Data*

Paddy Planting Area Data is used as input data in this research. The data used is data from three time periods: January to April, May to August and September to December, from 2009 to 2015. This data was obtained from Kalimantan Province’s Agriculture Department’s Agency of Crops and Horticulture as well as the Agency of Farming and Agriculture. This data is given the variable initial X2.

*c) Productivity Data.*

Productivity Data is used as input data in this research. The data used is data from three time periods: January to April, May to August and September to December, from 2009 to 2015. This data was obtained from Kalimantan Province’s Agriculture Department’s Agency of Crops and Horticulture as well as the Agency of Farming and Agriculture. This data is given the variable initial X3.

*d) Rainfall Data.*

Rainfall Data curah hujan digunakan sebagai data masukan pada penelitian. The data used is data from three time periods: January to April, May to August and September to December, from 2009 to 2015. This data was obtained from Kalimantan Province’s Agriculture Department’s Agency of Crops and Horticulture as well as the Agency of Farming and Agriculture. This data is given the variable initial X4.

*e) Rice Production Target Data*

Rice Production Target Data is used as output data in this research. The used data is from the year before the targeted year, so if the inputted data is from 2009, the rice production target data produced would be for 2010, and so on. This data was obtained from Kalimantan Province’s Agriculture Department’s Agency of Crops and Horticulture as well as the Agency of Farming and Agriculture. This data is given the variable initial Y.

*2) Neural Fuzzy System*

Neural Fuzzy System (NFS) was designed to realize fuzzy reasoning, where loads connected to the network are related to the parameters of fuzzy reasoning. By using a backpropagation learning algorithm, NFS could identify rules of fuzzy and train membership functions from the fuzzy reasoning. In NFS, the designing of the artificial neural network is guided by a formalization of fuzzy logic. It was then used to implement fuzzy logic, fuzzy decision-making, and also to represent membership functions that in turn represents fuzzy sets. [5].

*3) Calculating Predicted Value using Regression Interval with Neuro-Fuzzy*

Interval regression method with neuro-fuzzy is used to help people in making decisions. This method will explain the application of the design flow which has been analyzed and designed as a prediction method. Two processes are involved in the calculations, which are:

1. Interval regression with Neuro-Fuzzy process.
2. Prediction accuracy calculation process.

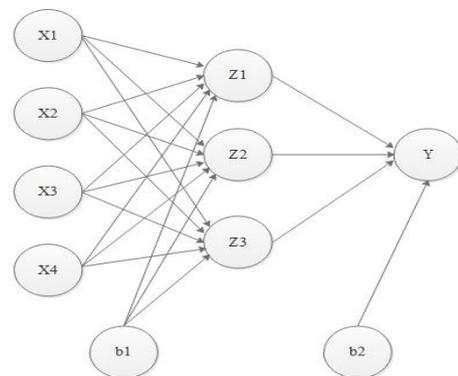


Fig. 2. Demonstrates the architecture of the interval regression with neuro-fuzzy, with 1 Hidden Layer and 3 Hidden Layer Cells

The regression model uses neuro-fuzzy which is a fuzzy interval regression model. For example, for a given input-

output pair  $(Xk, dk), k = 1, 2, \dots, p$  with  $Xk = (Xk1, Xk2, \dots, Xkn)$ . Fuzzy regression model on k-pattern is represented as (Lin, 1996):

$$Y(Xk) = A0 + A1Xk1 + \dots + AnXkn$$

With  $Ai$  being a fuzzy value. Because of that, the estimated output value of  $Y(xk)$  is also a fuzzy value. Fuzzy regression analysis can be simplified into regression interval analysis, from which the interval regression model will later be made.

The base concept of the interval regression analysis is developed from backpropagation networks, which are introduced by Ishibuchi and Tanaka in 1992. The model used 2 independently trained backpropagation network, with one used for the data interval upper limit, and the other used for the data interval lower limit. For example, if  $g^+(xk)$  and  $g^-(xk)$  are outputs from the two backpropagation networks (BPN<sup>+</sup> and BPN<sup>-</sup>) which are connected with the vector input  $xk$ , where every network has "n" neurons on the output layer. Learning process is carried out on the two networks (BPN<sup>+</sup> and BPN<sup>-</sup>) to produce network output  $g^+(xk)$  and  $g^-(xk)$  which are related with the following conditions (Lin, 1996):

$$g^-(x) \leq dk \leq g^+(x), k = 1, 2, \dots, p. \quad (1)$$

On the BPN<sup>+</sup> learning process, the used cost function was:

$$E = \sum_{k=1}^p E_k = \frac{1}{2} \sum_{k=1}^p \alpha_k [d_k - g^+(X_k)]^2 \quad (2)$$

With  $\alpha_k$  given as follows:

$$\alpha_k = \begin{cases} 1; & d_k >> g^+(x_k) \\ \alpha; & d_k \leq g^+(x_k) \end{cases} \quad (3)$$

Where  $\alpha$  is a small enough positive value in the interval (0, 1). This  $\alpha$  value can be given with a derivative function as:

$$\alpha(t) = \frac{1}{(1 + (\frac{t}{10000})^2)} \quad (4)$$

Where  $t$  is the value of iterations produced.

Backpropagation learning organization with three layers (1 input layer, 1 hidden layer, and 1 output layer) and a sigmoid activation function is used to generate  $w_j$  loads (the load between the hidden layer and the output layer) and  $w_{ji}$  (the load between the input layer and the hidden layer) on the BPN<sup>+</sup> network, with load rectification done as follows.

$$\Delta w_j = \eta \left( -\frac{\partial E_k}{\partial w_j} \right) = \eta \delta_k \quad (5)$$

$$\Delta w_{ji} = \eta \left( -\frac{\partial E_k}{\partial w_{ji}} \right) = \eta \delta_k y_{ki} \quad (6)$$

With

$$\delta_k = \alpha_k (d_k - y_k) y_k (1 - y_k) \quad (7)$$

$$\delta_{kj} = y_{kj} (1 - y_{kj}) \delta_k w_j \quad (8)$$

$y_k = g^+(x_k)$  is a network output and  $y_{kj}$  is the  $j$ -neuron output on the hidden layer for input  $x_k$ .

Following a similar logic, learning can also be done on the BPN<sup>-</sup> network to produce network output  $g^-(x_k)$ . The cost function used can be seen on equation (2.26), with  $\alpha_k$  given as follows:

$$\alpha_k = \begin{cases} \alpha; & d_k \geq g^-(x_k) \\ 1; & d_k < g^-(x_k) \end{cases} \quad (9)$$

Where  $\alpha$  is a small enough positive value in the interval (0, 1).

By using the two learning algorithms, it is possible to determine two functions— $g^+(x)$  and  $g^-(x)$ —where  $g^-(x) \leq dk \leq g^+(x), k = 1, 2, \dots, p$ . From here, an interval can be produced (Lin, 1996):

$$G(X) = [g^-(x), g^+(x)] \quad (10)$$

Because  $g^+(x)$  and  $g^-(x)$  is acquired through separate learning processes, it is very likely that  $g^+(x) < g^-(x)$ , in which case the interval could be modified as follows:

$$h^-(x) = \begin{cases} g^-(x); & g^-(x_k) \leq g^+(x) \\ \frac{1}{2} (g^-(x) + g^+(x)); & g^-(x_k) > g^+(x) \end{cases} \quad (11)$$

$$h^+(x) = \begin{cases} g^+(x); & g^-(x_k) \leq g^+(x) \\ \frac{1}{2} (g^-(x) + g^+(x)); & g^-(x_k) > g^+(x) \end{cases} \quad (12)$$

Therefore, the resulting interval is:

$$G(X) = [h^-(x), h^+(x)] \quad (13)$$

Description :

$Xk$  : Input Signal

$g^+$  : Data Interval Upper Limit

$g^-$  : Data Interval Lower Limit

$\alpha$  : Rate of Learning

$w_j$  : Load between hidden layer and output layer

$w_{ji}$  : Load between hidden layer and input layer

$\Delta w_j$  : Load correction between hidden layer and output layer

$\Delta w_{ji}$  : Load correction between hidden layer and input layer

BPN<sup>+</sup> : backpropagation+ results

BPN<sup>-</sup> : backpropagation- results

#### 4) Calculating Prediction Using Madaline Neural Network

Neural Network. Its interface was programmed using Visual Basic 2010, with Basic programming language. Two processes are at work during the calculation, which are:

a) *Madaline Neural Network Process.*

b) *Prediction Accuracy Calculation Process.*

The Madaline Neural Network Process consists of two stages: a training phase and a testing phase. The accuracy calculation process uses a Mean Square Error (MSE) value parameter. MSE is the squared average of the error difference between the testing phase output value compared to each targeted value. Figure 3. demonstrates Madaline Neural Network architecture with 1 Hidden Layer and 3 Hidden Layer Cells.

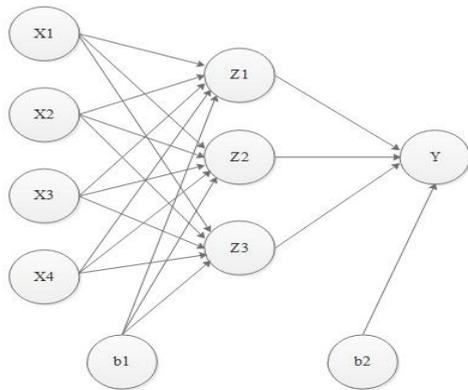


Fig. 3. Madaline Neural Network Architecture with 1 Hidden Layer and 3 Hidden Layer Cells

The Madaline Neural Network predictive algorithm initial pattern and bipolar target are as follows:

- a. Initialize all load and bias with small random values. Initialize learning rate ( $= \alpha$ ) with small values.
- b. While load shift is larger than tolerance or the number of epochs have not yet exceeded the predetermined limit, execute step 1 to 5.

1. Set unit activation input:  $x_i = s_i$  for all  $i$ .

2. Calculate net input for every hidden ADALINE unit

$$(z_1, z_2, z_3)$$

$$z_{in_j} = b_j + \sum x_i w_{ji} \quad (14)$$

3. Calculate the output of every hidden unit using the bipolar activation function:

$$z_j = f(z_{in_j}) = \frac{1}{1 + e^{-x}} \quad (15)$$

4. Determine network output:

$$y_{in} = b_k + \sum z_j v_j \quad (16)$$

$$y = f(y_{in}) = \frac{1}{1 + e^{-x}} \quad (17)$$

5. Calculate margin of error and determine load shift.

If  $y = \text{target}$ , then load shift will not be executed.

If  $y \neq \text{target}$  :

For  $t = 1$ , change load to the  $z_j$  dan  $z_{in}$  unit nearest to 0.

(For example, unit  $z_p$ ) as follows:

$$b_{p\text{new}} = b_{p\text{old}} + \alpha (1 - z_{in_p}) \quad (18)$$

$$w_{pi\text{ new}} = w_{pi\text{ old}} + \alpha (1 - z_{in_p}) x_i \quad (19)$$

For  $t = -1$ , change all loads to its positive  $z_k$  dan  $z_{in}$  unit, as follows:

$$b_{k\text{new}} = b_{k\text{old}} + \alpha (-1 - z_{in_k}) \quad (20)$$

$$w_{ki\text{ new}} = w_{ki\text{ old}} + \alpha (-1 - z_{in_k}) x_i \quad (21)$$

### III. RESULT AND ANALYSIS

#### A. Training Results

The research was conducted to predict rice production every three time periods using Interval Regression with Neural Fuzzy Method and Madaline Neural Network at Kubu Raya Regency to determine the interval upper boundary (BPN+) and interval lower boundary (BPN-) by using backpropagation algorithm during the training process. The rate of learning and testing was changed ten times for each three time periods.. After testing and changing the learning rate ten times for each three time periods, the best training results were gained after using a learning rate parameter value of 0.09. The tests using the Interval Regression with Neuro-Fuzzy Method and the Madaline Neural Network Method can be seen in Table 2. The best result was the one with the smallest MSE value, which in this case was 0.025324.

TABLE II. THE MSE VALUE OF TRAINING USING INTERVAL REGRESSION METHOD WITH NEURO-FUZZY AND TRAINING USING MADALINE NEURAL NETWORK METHOD.

No. Data	Period	Regression MSE Value	Madaline MSE Value
1.	January– April	0,064722 (BPN-)	0,05944
2.		0,030644 (BPN+)	
3.	May – August	0,141674 (BPN-)	0,25654
4.		0,179612 (BPN+)	
5.	September – December	0,025324 (BPN-)	0,32726
6.		0,036961 (BPN+)	

From the results of the training, it was concluded that the number of practice data has noticeable effect on the practice result. The more practice data used, the better the result. Smaller MSE values mean higher accuracy from training results. The epoch's progress was determined by the rate of learning used, with faster learning rate equating to less epoch used. If the rate of learning is too fast, however, it will disrupt the already established pattern into disarray.

#### B. Testing Results

In the testing process, the loads from the training result with the best network that produces the highest accuracy will be used to process the testing of the test data of three time

periods from 2014-2015. The prediction nears the estimated target and becomes more accurate when the value of the testing's error difference is smaller. Comparison of the Interval Regression with Neuro-Fuzzy Method with the Madaline Neural Network can be seen on Table 7. Based on the results of MSE calculations during the testing phase, it can be concluded that the best result with the smallest MSE value gained was 0,01098377.

TABLE III. TESTING MSE VALUE USING INTERVAL REGRESSION WITH NEURO-FUZZY METHOD AND THE MADALINE NEURAL NETWORK METHOD PENGUJIAN.

Interval Boundary	Regression MSE Value	Madaline MSE Value
BPN-	0,01098377	0,763362
BPN+	0,24760197	

#### IV. CONCLUSION

The factors that affect rice production are the size of the planting area, the harvesting area size, productivity, and rainfall. The best training results with the smallest MSE value of 0,01098377, was achieved with a 0.09 learning rate. The best MSE calculation results during the testing phase yielded the the smallest MSE value of 0,01098377. Results from both versions of the experiment suggest that the Interval Regression with Neuro-Fuzzy is the most appropriate model for the given task

#### ACKNOWLEDGMENT

The authors thank to the Faculty of Mathematics and Science, YSU, due to the funding of DIPA-2017 for this work.

#### REFERENCES

- [1] BPS Catalogue 510601861. West Kalimantan Province Figures from ST2013 Survey Result – Domestic Paddy Business 2013. BOOK A Rice Plants. (Angka Provinsi Kalimantan Barat Hasil Survei ST2013 – Subsektor Rumah Tangga Usaha Tanaman Padi 2014. BUKU A Tanaman Padi).
- [2] Rismawan, Tedy dand Kusumadewi, Sri. 2008. "Decision Support System for Interval Values of Body Fat Levels Using Interval Regression with Neural Fuzzy". ("Sistem Pendukung Keputusan Penentu Nilai Interval Kadar Lemak Tubuh Menggunakan Regresi Interval dengan Neural Fuzzy"). National Seminar of Teknoin 2008. Yogyakarta.
- [3] Siang, J.J. (2009.) "Artificial Neural Network and Programming it using MATLAB" ("Jaringan Syaraf Tiruan dan Pemogramannya menggunakan MATLAB").Yogyakarta : Publisher Andi (Penerbit Andi).
- [4] Supriyanto, 2012. "Broad Predictions of Harvesting Area and Rice Production at Banyumas Regency using Adaptive Neuro Fuzzy Inference System Method" ("Prediksi Luas Panen dan Produksi padi di kabupaten Banyumas menggunakan mteode Adaptive neuro fuzzy Inference System") :Purwokerto.
- [5] Pal, S. K; dan Ghosh, A. 1996. Neuro Fuzzy Computing For Image Processing and Pattern Recognition. International Journal of System Science vol 27 no 12.
- [6] Ying-Feng Wang, Xiao-Qin Zeng and Yeung. 2013. "Analysis of sensitivity behavior of Madalines", *Proceedings of 2005 International Conference on Machine Learning and Cybernetics*, Aug. 2005
- [7] Yingfeng Wang, Xiaoqin Zeng, Lixin Han, "Sensitivity of Madalines to Input and Weight Perturbations", *Proceedings of IEEE International Conference on Machine Learning and Cybernetics*, pp. 1349-1354, Nov. 2003.
- [8] KangaraniFarahani, Mahsa and Mehralian, Soheil."Comparison between Artificial Neural Network and neuro-fuzzy for gold price prediction". 13th Iranian Conference on Fuzzy Systems (IFSC), Aug 2013.