



## Graduate Prediction in University using Fuzzy Logic

Ainul Yaqin Ibnu Abdul Wahid<sup>1</sup>, Abba Suganda Girsang<sup>2</sup>

<sup>1</sup>Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, [ainul.wahid@binus.ac.id](mailto:ainul.wahid@binus.ac.id)

<sup>2</sup>Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, [agirsang@binus.edu](mailto:agirsang@binus.edu)

### ABSTRACT

STIMIK Sepuluh Nopember Jayapura is an educational organization engaged in technology and information. Student graduation rate from 2011 - 2018 seen from the number of incoming students has decreased. These conditions become reference material for the majors to conduct evaluations and strategic steps in an effort to increase the level of graduation of students so that quality objectives can be achieved. Evaluation of majors needs to be improved by analyzing the factors that cause late graduation for a student. 3. For the purposes of training data Student data is used from 2011 - 2015, while for testing data Student data is used for 2016. The method used to predict is the Adaptive Neuro Fuzzy Inference System method. Based on the test results, training is conducted on the amount of training data as much as 50%, 60% and 70% of the total data available. While testing is conducted on the remaining training data consisting of percentages of 50%, 40% and 30%. For each data combination ten times (Epoch value) and then the average RMSE value is taken from the test. The smallest RMSE value was found in a combination of 60% training data and 40% test data which was 0.006380. The more training data shows the higher the level of accuracy.

**Key words:** prediction, graduation, school, fuzzy logic

### 1. INTRODUCTION

For the world of education and higher education, in the era of globalization with the growth of information disclosure, political, economic, social, cultural, scientific and technological change is accelerating and becoming a very complex and interrelated challenge. In facing these global challenges, Universities need to prepare themselves well to face local and national demands. This condition is certainly realized by producing graduates who are able to compete given the increasingly fierce competition in gaining employment. When the College produces many graduates, it can automatically influence the improvement of reputation, promote quality, improve competitiveness, improve quality and of course influence the accreditation of the Higher Education.

STIMIK Sepuluh Nopember Jayapura is an educational organization which is engaged in technology and information. Student graduation rate from 2011 - 2018 seen from the number of incoming students has decreased. These conditions become reference material for the majors to conduct evaluations and strategic steps in an effort to increase the level of graduation of students so that quality objectives can be achieved. Evaluation of majors needs to be improved by analyzing the factors that cause late graduation for a student. In general, the duration of study of students (Strata 1) is within a period of 4 (four) years and is the fastest within 3.5 (3 and a half) years. However, not all students can complete their studies at the allotted time. It is reasonable if the prediction of the length of study of students is needed by the management of Higher Education and can be a convenient service for students.

Fuzzy logic is one of the decision support systems to assist in the decision making process that can broaden the capabilities of decision makers without changing their judgment in making a decision[1]. In a study conducted by [2] the application of the Artificial Neural Network algorithm (ANN) to predict student graduation with outputs of passing late, fast and not on time obtained good accuracy results. ANN has the ability to learn, but can not explain the reasoning process that is done because the knowledge it has is only in the form of weights that are of real value, whereas Fuzzy Logic does not have the ability to learn, but can explain the reasoning process that is done based on the rules on the basis the knowledge he has. The two methods have opposite characteristics but when combined they will become a better method. In research conducted by [3] the combination of Fuzzy Logic and ANN or commonly called ANFIS (Adaptive Neuro Fuzzy Inference System) produces a more optimal value. ANFIS is a concept between Backpropagation Neural Network and Fuzzy logic concept. Fuzzy rules are used for optimization, while ANN is used to produce optimum fuzzy rules for use in the fuzzification process.

In this research, ANFIS algorithm is used. This algorithm has the ability to learn in the training process based on information obtained from linguistic variables or processes from the Fuzzy inference system. The features used in this study are Temporary Achievement Index semester 1 to 4, number of SKS graduated semester 4, parents' income, entrance test scores, distance of residence and age. Using this method is expected to help and facilitate STIMIK Sepuluh

Nopember Jayapura in predicting student graduation, so as to minimize the number of students who are late or not graduating in the future and can also be used as an early warning to Lecturer Guardians or Heads of Study Programs on the condition of student study performance.

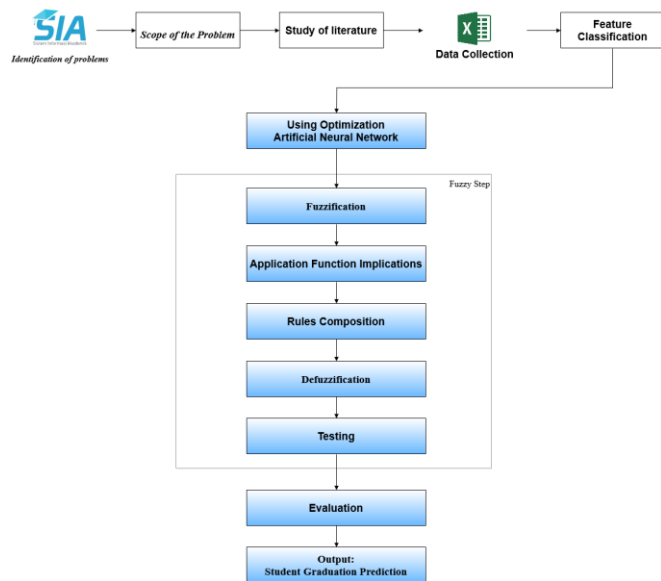
**1.1 References**

Prediction of Student Graduation with Artificial Neural Networks[2]. Use of Artificial Neural Networks Backpropagation Method. Input variables: Name, Place of Birth, Date of Birth, Age, Number of SKS, average semester GPA 1-4 Output variables: late graduation, fast and not on time. The optimization method is not used, so the membership function is created manually with estimated values.

Prediction of Student Graduation Using Neural Network Method and Particle Swarm Optimization[4]. Use of Neural Network Methods and Particle Swarm Optimization. Input variables: Population, RMSE, Epoch and Time. Output variable: Pass on time and pass on time. Cannot explain the reasoning process that is done based on the rules (rules) in the knowledge base they have.

The Effect of GPA and Motivation in Predicting the Accuracy of Student Graduation Times by Using an Adaptive Neuro Fuzzy-Based Expert System[3]. Use of the Adaptive Neuro Fuzzy Inference System Method. Input variables: GPA and Motivation. Output variable: Pass on time and pass on time. Only 2 features are used, it needs additional features to get better results.

**1.2 Framework of Thinking**



**Figure 1:** Framework of Thinking

As explained in the previous chapters, the ANFIS method is used in this study. The membership function in this study uses a membership function related to student graduation at STIMIK Sepuluh Nopember Jayapura whose membership values are optimized. So the membership value in each membership function is not determined manually, but the

membership function is obtained automatically. After obtaining the membership function and the membership value, fuzzification is then performed using the Tsukamoto method, which then obtained the value of the results of defuzzification that can help decision makers to see whether students whose data is used in fuzzification can graduate on time or not. The framework of thinking can be shown in Figure 1.

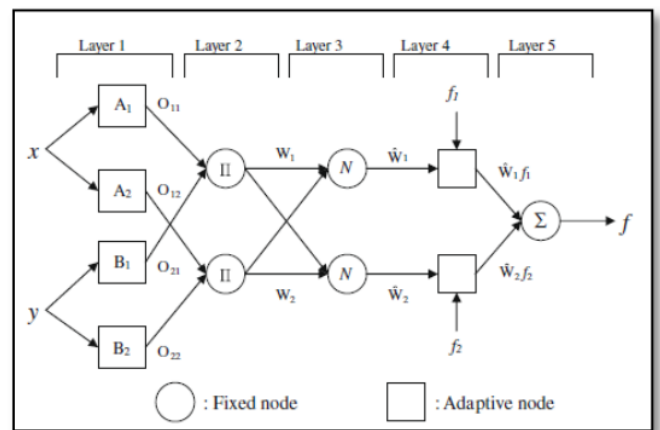
**2. DATA**

In this case used input data consisting of five variables, namely GPA, parents' income, age, entrance test scores, and distance (KM). The data to be used in the model trials are 421 data which will be divided into training data and test data. From 421 sample data, 70% will be used as training and the rest will be used as test data.

**2.1 ANFIS**

Adaptive Neuro Fuzzy Inference System (ANFIS) is a combination of fuzzy logic and artificial neural networks (ANN). Fuzzy logic has advantages in modeling the qualitative aspects of human knowledge and the decision making process by applying a rule base (rules). ANN has advantages in recognizing patterns, learning and practicing in solving problems without requiring mathematical modeling. And can work based on historical data entered into it and can predict future events based on these data. So that ANFIS has both capabilities [5].

The framework of the ANFIS method has five layers, namely the fuzzification layer, the rule layer, the normalization layer, the defuzzification layer, and the single neuro result. The ANFIS layers is shown in Figure 2.



**Figure 2:** ANFIS Layers

Anfis structural model used in this study uses five features. The ANFIS model structure is shown in Figure. 3

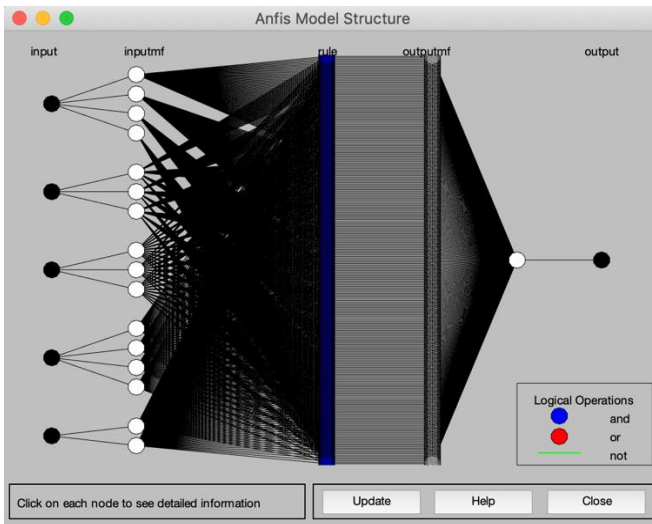


Figure 3: ANFIS Model Structure

Display surface viewer to see the entire output surface of the system, i.e. the overall span of the output set based on the overall span of the input set. In this picture a two-dimensional curve is presented that presents a mapping from one or two inputs to the output. The surface viewer structure is shown in Figure 4.

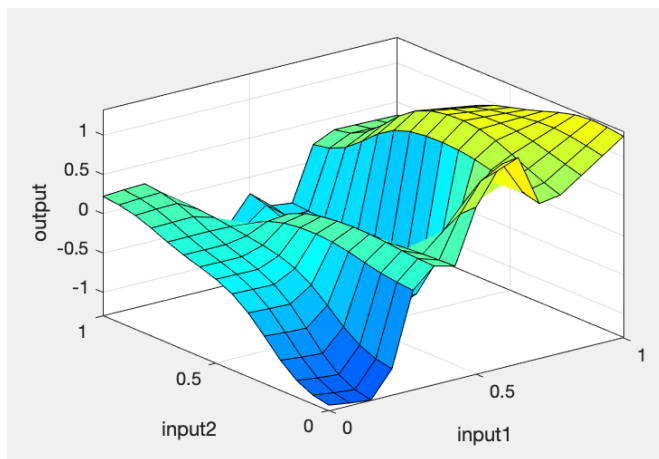


Figure 4: Surface Viewer

## 2.2 ANFIS Training

For FIS training, ANFIS provides two methods of parameter optimization for membership functions, namely Backpropagation and Hybrid (combined backpropagation and least square). This time we will try to use Hybrid optimization. To stop the training process, ANFIS uses the value of error tolerance, so that if after training the error data enters this error tolerance area then the training will stop. An error tolerance of 0 (default) is selected. Then the number of epochs (iterations) of the training process is determined by 10. The results is shown in Figure 5.

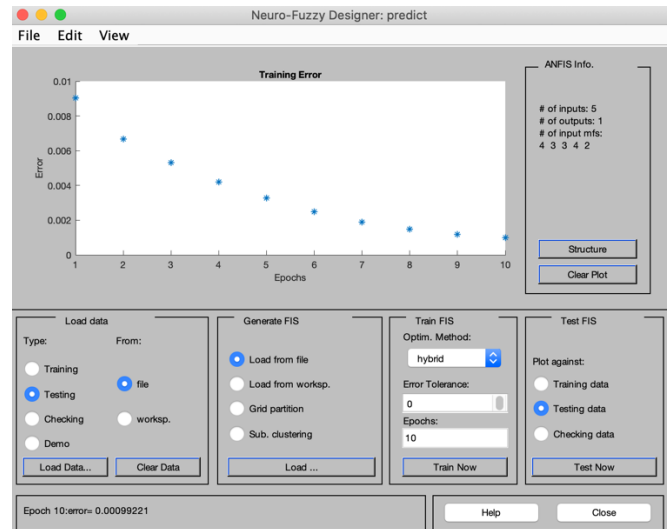


Figure 5: ANFIS Training

The training error graph decreases and reaches a fixed error point at 0.0009221 when the epoch reaches 10. This means that the training process produces a minimum error at the value of 0.000866, which is the difference of 0.00012621 from the error tolerance that we define (never reaches the value of 0), for the training process is limited by the number of iterations (epoch), which is 10 times.

## 2.3 Testing

To find out the ANFIS performance that has been trained with preliminary data (trainData), we can do a test by plotting trainData and ANFIS data with the same input values. In the FIS Test column in the ANFIS editor click Test Now by previously selecting training data. The ANFIS testing is shown in Figure 6.

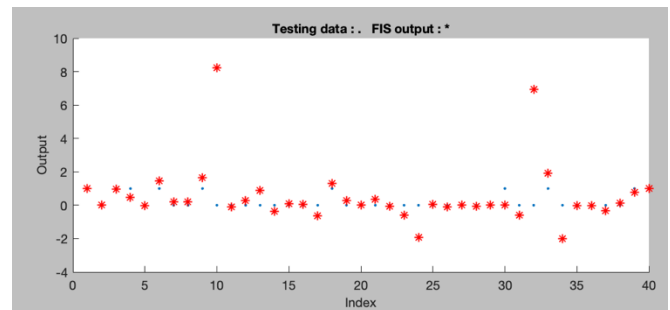


Figure 6: ANFIS Testing

## 3. ANALYSIS OF ANFIS METHOD

After normalizing the data, then we will look for inference values using the ANFIS method. Calculation of the ANFIS method will be calculated at each layer. Fuzzy inference system applied is Takagi Sugeno Kang's fuzzy inference.

### 3.1 First Layer

In the first layer, the fuzzification process occurs. This process is useful for mapping data input into fuzzy sets. In this

process a fuzzy membership function calculation will be performed to transform the classical set input to a certain degree using 20 sample data. The membership function used is the Generalized-Bell type in which the membership function contains two parameters, the mean and the standard deviation. Results of clustering with the FCM algorithm can be shown in Table 1.

**Table 1:** Clustering Algorithm FCM

No	Membership Degrees					Cluster tendencies				
	X1	X2	X3	X4	X5	1	2	3	4	5
1	0,872000	0,333333	0,000000	0,857143	0,842105	*				
2	0,920000	0,666667	0,125000	0,771429	0,473684	*				
3	0,700000	0,000000	0,125000	0,571429	0,210526	*				
4	0,600000	0,333333	0,625000	0,285714	0,736842					*
5	0,640000	0,333333	0,312500	0,428571	0,526316	*				

For example the Generalized Bell (gbellmf) membership function is used in Eq. (1)

$$\mu_A(x) = \frac{1}{1 + \left| \frac{x - c}{a} \right|^{2b}} \tag{1}$$

Determined b = 1 and as an initialization of the values a and c, the mean and standard deviation are used. Result of Mean and Standard Deviation can be shown in Table 2.

**Table 2:** Values of Mean and Standard Deviation

<i>mean</i>	0,806286	0,300000	0,178125	0,534921	0,393352
	0,531385	0,666667	0,178125	0,857143	0,736842
<i>sd</i>	0,000000	0,010541	0,077564	0,082683	0,006355
	0,000000	1,159502	0,077564	0,455691	0,343490

Then neurons from layer 1 are counted is used in Eq. (2)

$$\mu_{A1} = \frac{1}{1 + \left| \frac{x1 - 0,627600}{0,267496} \right|} \tag{2}$$

Layer one output is the degree of membership of each data, the result is can be shown in Table 3.

**Table 3:** First layer results

Data	Membership Degrees									
	A1	A2	B1	B2	C1	C2	D1	D2	E1	E2
1	4,07758E-31	7,35514E-32	0,090910259	0,199834382	0,159391272	0,159391272	0,06177723	1	0,000200507	0,914149525
2	1,36173E-31	5,65041E-32	0,000825775	0,02458701	0,680682823	0,680682823	0,108909143	0,965828277	0,00621931	0,630137882
3	1,55874E-31	3,00141E-31	0,001233063	0,159391272	0,680682823	0,680682823	0,836851893	0,717814045	0,001206792	0,29870202
4	4,13794E-32	1,81249E-30	0,090910259	0,199834382	0,029245411	0,029245411	0,099164961	0,388730988	0,00034218	1
5	6,36814E-32	7,23331E-31	0,090910259	0,199834382	0,24991569	0,24991569	0,376733774	0,530640157	0,002279153	0,726928742

### 3.2 Second Layer

The output of this layer is the result of multiplication of all incoming signals. Each node output states the degree of activation of the fuzzy rules is used in Eq. (3). Results of the second layer can be shown in Table 4.

$$w_1 = \mu_{A_1}(x) \cdot \mu_{B_1}(x) \quad ; \quad w_2 = \mu_{A_2}(x) \cdot \mu_{B_2}(x) \tag{3}$$

**Table 4:** Second layer results

Data	Second Layer Result 2	
	W1	W2
1	7,31876E-38	2,14162E-33
2	5,18449E-38	5,75527E-34
3	1,32125E-37	6,98208E-33
4	3,73308E-39	4,11769E-33
5	1,2423E-36	1,39345E-32

### 3.3 Third Layer

Each node in this layer is a nonadaptive node that displays the normalized firing strenght function, which divides wi by the total number of w is used in Eq. (4). Results of the third layer can be shown in Table 5.

$$\bar{w}_1 = \frac{w_1}{w_1 + w_2} \quad ; \quad \bar{w}_2 = \frac{w_2}{w_1 + w_2} \tag{4}$$

**Table 5:** Third layer results

Data	Third Layer 3	
	W1	W2
1	3,41727E-05	0,999965827
2	9,00743E-05	0,999909926
3	1,8923E-05	0,999981077
4	9,06594E-07	0,99999093
5	8,91447E-05	0,999910855

### 3.4 Fourth Layer (Defuzzification)

**Table 6:** Fourth layer results

Data	p	q	r	s	t	u
1	2,97986E-05	1,13909E-05	3,41727E-05	0	0,857113566	0,999965827
2	3,14389E-05	2,27818E-05	9,00743E-05	0,124995728	0,771402209	0,999909926
3	2,39209E-05	0	1,8923E-05	0,124995728	0,571409044	0,999981077
4	2,05036E-05	1,13909E-05	9,06594E-07	0,624978642	0,285704522	0,999999093
5	2,18705E-05	1,13909E-05	8,91447E-05	0,312489321	0,428556783	0,999910855

At this layer, calculations are made to convert the fuzzy results into a classical set (crisp). At this layer the LSE calculation is performed to get the coefficient parameter values. Results of the fourth layer can be shown in Table 6.

### 3.5 Fifth Layer

At this layer there is only one fixed node whose function is to add up all the input coming from the fourth layer. Results of the fifth layer can be shown in Table 7.

**Table 7:** Fifth layer result

Data	Z target	P	Z output	Error
1	1	1,857154755	1	0,857154755
2	1	1,896452158	1	0,896452158
3	1	1,696428693	1	0,696428693
4	1	1,910715058	1	0,910715058
5	1	1,741079365	1	0,741079365

## 4. RESULTS

For RMSE testing, training is carried out on 50%, 60% and 70% of the total training data. While testing is conducted on the remaining training data consisting of percentages of 50%, 40% and 30%. For each data combination ten times (Epoch value) and then the average RMSE value is taken from the test. Results of this research can be shown in Table 8 and 9.

**Table 8:** Results

Data	IPK	Income	Age	Test scores	distance of residence	Status	Value	Result
1	3,55	5000000	21	90	3	1	1.4163	1
2	3	5000000	25	60	1	0	0.5349	0
3	3,25	7000000	24	85	1	1	-0.6015	0
4	3,7	8000000	24	70	5	1	0.6334	0
5	2,4	5000000	24	75	7	0	-1.6980	0

**Table 9 :** The results of testing the RMSE value in the prediction process

Percent of Train Data	Percent Test Data	RMSE
50%	50%	0.046089
60%	40%	<b>0.006380</b>
70%	30%	0.046089

Based on table 9 it can be seen that the smallest RMSE value is found in a combination of 60% training data and 40% test data. The more training data shows the higher the level of accuracy. However, this is not the case with the RMSE value, which is different in every combination of training and test data.

## 5. CONCLUSION

Based on the problem identification in the discussion of the previous chapter, the following conclusions can be obtained: The results of the analysis of the ANFIS method show there is still an excess of predictive value error. This is due to lack of data used.

The combination of % data and Epoch value on ANFIS can affect the accuracy of predictions based on RMSE values. The more training data, the more accurate the predicted value obtained.

## REFERENCES

1. Haryanto. (2016). Optimalisasi Penentuan Kelayakan Kredit Nasabah Pada CU Khatulistiwa Bhakti Pontianak Menggunakan Fuzzy Logic. Binus University Jakarta.
2. Ansari, R. (2013). Syaraf\_Tiruan\_Propagation4. Prediksi Kelulusan Mahasiswa Dengan Jaringan Syaraf Tiruan, JTIULM Vol 1, No 1 Jan-Jun 2016: 18-23
3. Welim, Y. Y., Informasi, F. T., & Luhur, U. B. (2013). Menggunakan Sistem Pakar Berbasis Adaptive Neuro, 2–4.
4. Kusumawati, D., Winarno, W. W., & Arief, M. R. (2015). Prediksi kelulusan mahasiswa menggunakan metode neural network dan particle swarm optimization, 6–8. Semnas Teknologi Informasi dan Multimedia 2015. ISSN 2302-3805
5. Sitohang, S., Girsang, A.S., & Suharjito, S.(2017). Prediction of the number of airport passengers using fuzzy C-means and adaptive neuro fuzzy inference system. International Review of Automatic Control, 10(3), 280–287.
6. Ceylan, Zeynep & Pekel, Ebru & Ceylan, Selim & Bulkan, Serol. (2018). Biomass higher heating value prediction analysis by ANFIS, PSO-ANFIS and GA-ANFIS. Global Nest Journal. 20. 589-597. 10.30955/gnj.002772.
7. Munawir, Munawir & Iqbal, Taufiq. (2019). Prediksi Kelulusan Mahasiswa menggunakan Algoritma Naive Bayes (Studi Kasus 5 PTS di Banda Aceh). Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi). 3. 59. 10.35870/jtik.v3i2.77.
8. Jain, Dr Vineet & Raj, Tilak. (2018). Prediction of cutting force by using ANFIS. International Journal of System Assurance Engineering and Management. 9. 1-10. 10.1007/s13198-018-0717-x.