Model Adaptive Fuzzy Time Series to Forecasting Enrollments of New Student

Ica Admirani[1,*] Ikhthison Mekongga[1] Isnaini Azro[1] Hidayati Ami[1]
Rian Rahmanda Putra[1]

[1] *Department of Computer Engineering, State Poytechnic of Sriwijaya.*
[*]*Corresponding author. Email: ica_admirani@polsri.ac.id*

ABSTRACT
The estimated enrollments of new student is required in the academic planning of a higher education institution. That can be done by forecasting using the fuzzy time series (FTS) technique. FTS method is an artificial intelligence computation technique that can capture patterns from previous data to predict future event. The implementation of this method is easier to used. In this study, the Adaptive model to FTS is applied to forecast new student, where the interval division is done twice (twice-divided) and weighting is carried out for the prediction process. The prediction results obtained a deviation value (error) of 11.66% which is measured using MAPE. These results indicate that this model can be used for long-term predictions even with a limited sample of data.

Keywords: *Forecasting, Adaptive FTS, Interval, Twice-divided, weighting, MAPE*

1. INTRODUCTION

Students are very important in ta higher educational institusion, if the college does not have students or has few students then the learning activities in the college will not go well. To get students, every year the college opens a new student registration[1].

The growth in the number of students can affect the academic plan of a higher educational institusion. As happened in the Department Computer Engineering, Sriwijaya State Polytechnic where the growth in the number of students becomes a parameter of opportunities to open new study programs, adding classes to existing study programs, increasing the number of seats in the class and managing class schedules. This can be estimated by forecasting the number of new student growth.

Fuzzy time series is an artificial intelligence technique that can be used for forecasting. This technique can capture the pattern of the data in the past to predict events in the future based on these data [2]. Song and Chissom first introduced this method in 1993 to predict enrollments of new students data at the University of Alabama by applying the Time-Invariant FTS model. The model shows a relatively small deviation over time. This model that has good resistance even if the historical data is less accurate [3]. FTS have also been used to estimate the number of prospective new students in the Department of plantation crop cultivation at the Samarinda Agricultural Polytechnic. This is done because the forecasting method is time independent [4]. This method is also used to solve the problem of time series data which is non stationary and uncertain, such as the forecasting of bandwidth usage with fuzzy time series Singh's model[5], company profit with heuristic time invariant model [6] cement demand [7], the tourist visits [8] and the pollutant PM_{10} concentration with adaptive fuzzy time series[9]. In this research, AdaptiveFTS model was applied to predict the registration of new students at the Computer Engineering Department, Sriwijaya State Polytechnic.

2. LITERATURE REFERENCES

2.1 Fuzzy Logic Theory

Fuzzy logic theory was first introduced by Zadeh in 1965, which is a logic that has an obscure value between true or false. Fuzzy logic allows membership values to be between 0 and 1 [10]. Fuzzy logic has advantages such as; The concept is easy to understand, because in this theory there are simple mathematical concepts that underlie fuzzy reasoning; Very flexible, which is able to adapt to changes and uncertainties that accompany problems; Have tolerance for inappropriate data; Able to model complex nonlinear functions; Without going through the training process can directly apply the experience of the experts; Based

on natural language, namely using everyday language so that it is easy to understand; Can cooperate with conventional control techniques.

3. RESEARCH METHODOLOGY

3.1 Adaptive FTS Model

The Adaptive FTS model or also known as the Ruey Chyn Tsaur FTS model has the following prediction steps [8] [9]:

1. Establish a universe of discourse U from historical data.

2. The Universe of discourse U is divided into equal intervals.

3. The interval in step 2 is redistributed, if the number of linguistic values generated is still large compared to the average number, then the original linguistic values must be further divided into half of each. Because there are so many data in linguistic values, using twice-divided linguistic values to display data is more reliable than once-divided linguistic values.

4. Determine the fuzzy sets in universe U.

5. Fuzzyfication of historical data and creating Fuzzy Logical Relationship (FLR).

6. Create fuzzy logical relation group (FLRG) and flux type matrix. Example, if $\boldsymbol{A2 \rightarrow A1, A2 \rightarrow A2, A2 \rightarrow A1}$, and it can be explained that $\boldsymbol{A2 \rightarrow A1, A2, A1}$. All *FLRG* have a flux-type matrix. Estimates of the FLRG were carried out in succession and designed the weights into different trends as follows:
   (t=1) $A_2 \rightarrow A_1$ with weight of 1
   (t=2) $A_2 \rightarrow A_2$ with weight of 2
   (t=3) $A_2 \rightarrow A_1$ with weight of 3
   (t=4) $A_1 \rightarrow A_2$ with weight of 1
   FLR (t = 3) determines that the highest weight is 3, where the probability that the next midpoint will appear is higher than the others. On the other hand, FLR (t = 1) is determined that the weight is less than 1, where the probability that the next midpoint will appear is lower than the others. So that a fluctuation type matrix can be made for all FLR.

7. Determination of Weights. Because each FLR for each FTS data usually repeats r based on time ∀r≥1, we can classify it and provide weighting reasons according to frequency. The matrix from step 6 is the standard for Wn, then we can define the weight matrix as follows:

$$W_n(t) - [W'_1, W'_2, \ldots, W'_i] \quad (3.1)$$
$$\left[\frac{W1}{\sum_{k=1}^{i} Wk}, \frac{W2}{\sum_{k=1}^{i} Wk}, \ldots, \frac{Wk}{\sum_{k=1}^{i} Wk}\right]$$

8. Calculation of the predicted value. Based on step 5, the predictive value is obtained as follows:

$$F(t)=L_{df}(t-1)\mathrm{o}W_n(t\text{-}1)_1 \quad (3.2)$$

   Where L_{df} (t–1) is the defuzzyfication matrix and W_n(t–1) is the weighted matrix.

9. Testing the applied FTS methot is measured using Mean Absolute Percentage Error

$$\text{MAPE}=\frac{1}{n}\sum_{t=1}^{n}\left|\frac{\textit{Actual value }(t)-\textit{Forecast }(t)}{\textit{Actual value }(t)}\right| \quad (3.3)$$

3.2 Block Diagram of Adaptive FTS Model

The block diagram of Adaptive FTS model is showed on figure 1

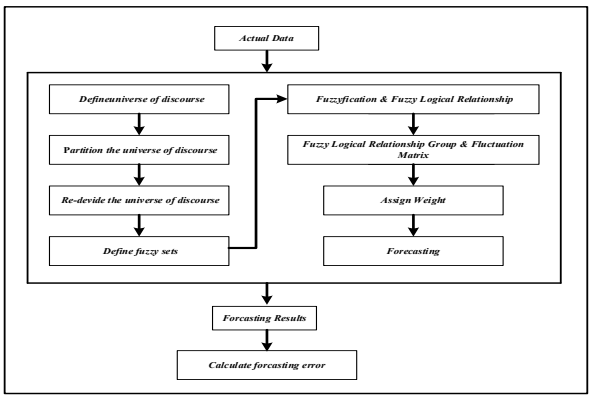

Figure 1. Block Diagram of Adaptive FTS Model

From Figure 1, it can be seen that the actual data in the form of historical data for new student registration is processed using the adaptive FTS model to predict student enrollment according to the respective stages. The output or result of the prediction process for student enrollment is then calculated the prediction error.

4. RESULTS AND DISCUSSION

4.1 The Results of Actual Data Collection

The actual data used in this study is historical data for the registration of new students at Computer Engineering Department, Sriwijaya State Polytechnic from 2004 to 2019 with 16 sample data as shown in table 1.

Table 1. Actual Data

Part	Years	Number of registrant
1	2004	219
2	2005	245
3	2006	220
4	2007	346
5	2008	444
6	2009	517
7	2010	624
8	2011	573
9	2012	586
10	2013	699
11	2014	699
12	2015	538
13	2016	867
14	2017	871
15	2018	1164
16	2019	1151

4.2 Forecasting Process For Adaptive FTS Model

The forecasting process using the adaptive FTS model, there are several steps that must be done, such as:

4.2.1 Determine The Universe Discourse

From the historical data in table 1, the lowest value (Dmin) is 220 and the highest value (Dmax) is 1164, then from the Dmin and Dmax values it is determined that D1 = 20 and D2 = 36 in order to obtain the value of the universe set (universe discourse) U = (200, 1200].

4.2.2 Determine The Interval

Furthermore, the universe of discourse (U) is divided into the same intervals (ui), namely u1 = [200,300], u2 = [300,400], u3 = [400,500], u4 = [500,600], u5 = [600,700], u6 = [700,800], u7 = [800,900], u8 = [900, 1000], u9 = [1000,1100] and u10 = [1100,1200]. From the results of this division, the amount of historical data included in each interval is obtained as shown in Table 2.

Table 2. The amount of historical data in intervals the universe of discourse

Part	Interval (u_i)	Amount
1	u1	3
2	u2	1
3	u3	1
4	u4	4
5	u5	3
6	u6	0
7	u7	2
8	u8	0
9	u9	0
10	u10	2

4.2.3 Interval redistribution

The division of the interval into half of the value of the interval to the number of historical data is greater than the number of intervals divided by two of the division of the first set of universes, namely 10/2 = 5.From the intervals in table 4.2, it can be seen that there is no number of members greater than 5, so the interval remains as much as 10 because no interval is subdivided. Next, determine the mean value of each interval, namely m1 = 250, m2 = 350, m3 = 450, m4 = 550, m5 = 650, m6 = 750, m7 = 850, m8 = 950, m9 = 1050 and m10 = 1150.

4.2.4 Determine Fuzzy Set

After determining the intervals of the speaking universe, then determining the fuzzy set, namely A1 = [200,300], A2 = [300,400], A3 = [400,500], A4 = [500,600], A5 = [600,700], A6 = [700,800], A7 = [800,900], A8 = [900,1000], A9 = [1000,1100] and A10 = [1100,1200].

4.2.5 Fuzzyfication of Historical Data and Create FLR

By assigning fuzzy sets A membership, fuzzification is carried out on historical data as shown in Table 3.

Table 3. Fuzzyfication of Historical Data

Part	Years	Actual Data	Fuzzifications
1	2004	219	A1
2	2005	245	A1
3	2006	220	A1
4	2007	346	A2
5	2008	444	A3
6	2009	517	A4
7	2010	624	A5
8	2011	573	A4
9	2012	586	A4
10	2013	699	A5
11	2014	699	A5
12	2015	538	A4
13	2016	867	A7
14	2017	871	A7
15	2018	1164	A10
16	2019	1151	A10

From the result of *FLR* are $A_1 \rightarrow A_1$, $A_1 \rightarrow A_1$, $A_1 \rightarrow A_2$, $A_2 \rightarrow A_3$, $A_3 \rightarrow A_4$, $A_4 \rightarrow A_5$, $A_5 \rightarrow A_4$, $A_4 \rightarrow A_4$, $A_4 \rightarrow A_5$, $A_5 \rightarrow A_5$, $A_5 \rightarrow A_4$, $A_4 \rightarrow A_7$, $A_7 \rightarrow A_7$, $A_7 \rightarrow A_{10}$, $A_{10} \rightarrow A_{10}$.

4.2.6 Assign Fuzzy Logical Relationship Groups (FLRG)

The next step is the formation of FLRG, namely:

Group $A_1 \rightarrow A_1, A_2$.
Group $A_2 \rightarrow A_3$
Group $A_3 \rightarrow A_4$
Group $A_4 \rightarrow A_4, A_5, A_7$.
Group $A_5 \rightarrow A_4, A_5$.
Group $A_7 \rightarrow A_7, A_{10}$.
Group $A_{10} \rightarrow A_{10}$.

4.2.7 Assign Weight Value

From the FLRG, a fluctuation-type matrix is built as shown in Table 4.

Table 4. *fluctuation-type matrix*

$P(t-1)$	$P(t)$									
	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
A_1	2	1	0	0	0	0	0	0	0	0
A_2	0	0	1	0	0	0	0	0	0	0
A_3	0	0	0	1	0	0	0	0	0	0
A_4	0	0	0	1	2	0	1	0	0	0
A_5	0	0	0	2	1	0	0	0	0	0
A_6	0	0	0	0	0	0	0	0	0	0
A_7	0	0	0	0	0	0	1	0	0	1
A_8	0	0	0	0	0	0	0	0	0	0
A_9	0	0	0	0	0	0	0	0	0	0
A_{10}	0	0	0	0	0	0	0	0	0	1

Furthermore, weighting is carried out according to the frequency of the fluctuation-type matrix, for example from Table 4.4 there is P (t-1) = A4, then the predictions for P (t) are A4, A5, and A7 from the weighting matrix are [1/4, 2/4, 1/4].

4.2.8 Forecasting Value Calculation

The forecast process is carried out for each actual data value. The following is the calculation for predicting student enrollment in 2008, then the calculations are as follows:

$F(2010) = L_{df}(2009)$ o $W_n(2009)$.
$F(2010)$ = [m4, m5, m7] o [1/4, 2/4, 1/4] = [550, 650, 850] o [1/4, 2/4, 1/4]

$F(2010) = 67$.

Where L_{df} (2009) is the result of the defuzzification matrix and W_n (2009) is the result of the weight matrix. The prediction results for actual student enrollment data from 2005 to 2019 with 16 sample data are shown in table 5

Tabel 5 The results of Forecasting

Part	Years	Actual Data	Prediction Value
1	2004	219	-
2	2005	245	283,3
3	2006	220	283,3
4	2007	346	283,3
5	2008	444	450
6	2009	517	550
7	2010	624	675
8	2011	573	583,3
9	2012	586	675
10	2013	699	675
11	2014	699	583,3
12	2015	538	583,3
13	2016	867	675
14	2017	871	1000
15	2018	1164	1000
16	2019	1151	1150
17	2020	?????	1150

4.2.9 Testing The Method

From the results of the forecasting of new student enrollment with this adaptive FTS model, then testing is carried out by calculating the value of deviation (error) on the forecasting results using the MAPE. In the test results obtained prediction error of 11.66%.

5. CONCLUSION

From the results of the forecasting of new student enrollment with an adaptive fuzzy time series model using 16 sample data, evaluated using MAPE. From the test results, the deviation value (error) was 11.66%. From the results of this study, it can be seen that this adaptive FTS can be used for long-term predictions even though it has a limited data sample.

REFERENCES

[1] Pamungkas, G.W., and Kardian, A.R., 2013, Analysis of Estimated Number of New Student Admissions Using The Smallest Squares Method and SPSS17.0 Study Case: STMIK JAKARTA (STI&K), *Jurnal Ilmiah Komputasi* Vol 12 Part. 2

[2] Robandi, I., 2006, *Modern Power System Design, Optimization, Fuzzy Logic and Algorithms Genetika,* Andi, Yogyakarta.

[3] Song, Q., dan Chissom, B., 1993, Forecasting Enrollments with Fuzzy Time Series part 1. *Fuzzy Sets and System* 54, 1-9.

[4] Karmita, Sisie, etc, 2018, Predicted Number of Prospective New Students Using Fuzzy Time Series-Time Invariant, Proceeding seminar on Computer Science and Information Technology, Vol. 3 Pat. 1

[5] Aryanti and Mekongga, I, 2015, The Application of Fuzzy Time Series Singh for Forcasting Banwidth Network Demand, Proceeding of International Conference on Electrical Engineering, Computer Science and Informatics (EECSI 2015).

[6] Admirani, I., 2018, Implementation of Fuzzy Time Series Method For Profit Prediction In Companies, Jupiter, Vol 10, Part. 1

[7] Evriyantino, Y, Setiawan, B.D. and Wihandika, R.C., 2019, Cement Demand Prediction With Fuzzy Time Series Method, Journal of Information Technology development and Computer Science, Vol. 3 Part. 9

[8] Tsaur, R.C., dan Kuo, T.C., 2011, "The Adaptive fuzzy time series model with an application to Taiwan's tourism demand "*International Journalof Expert System with Applications,* Vol. 38 (2011) 9164-9171

[9] Aripin, A., Suryono, S., dan Surarso, B., 2016, Web Based Prediction Of Pollutant Concentration Using Ruey Chyn Tsaur Fuzzy Time Sries Model, AIP Conference Proceedings 1746, 020046.

[10] Kusumadewi, S., dan Purnomo, H., 2013, *Fuzzy Logic Application for Decision Supporters Issue 2*, Graha Ilmu, Yogyakarta.

[11] Admirani, I., Gernowo, R., and Suryono, S., 2016, Heuristic Time Invariant Fuzzy Time Series Model And Regression For Profit Prediction And Influencing Variable Analysis, Journal of Business Information Systems, Vol 6, part. 2.