

Fuzzy Time Series for Projecting School Enrolment in Malaysia

Nor Hayati Shafii^{1*}, Rohana Alias², Siti Rohani Shamsudin³, Diana Sirmayunie Md Nasir⁴

^{1,2,4} Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA, Perlis Branch, Arau Campus, 02600 Arau, Perlis, Malaysia

Corresponding author: *norhayatishafii@uitm.edu.my

Received Date: 18 January 2021

Accepted Date: 7 March 2021

Published Date: 14 March 2021

HIGHLIGHTS

- Fuzzy Time Series is used to develop models to forecast the number of school enrolments in Malaysia.
- Model is developed for the enrolment of pre-primary, primary, secondary, and tertiary school.
- The accuracy of the model is tested based on the value of mean squared error (MSE), Mean Absolute Percent Error (MAPE) and Mean Absolute Deviation (MAD).

ABSTRACT

There are a variety of approaches to the problem of predicting educational enrolment. However, none of them can be used when the historical data are linguistic values. Fuzzy time series is an efficient and effective tool to deal with such problems. In this paper, the forecast of the enrolment of pre-primary, primary, secondary, and tertiary schools in Malaysia is carried out using fuzzy time series approaches. A fuzzy time series model is developed using historical dataset collected from the United Nations Educational, Scientific, and Cultural Organization (UNESCO) from the year 1981 to 2018. A complete procedure is proposed which includes: fuzzifying the historical dataset, developing a fuzzy time series model, and calculating and interpreting the outputs. The accuracy of the model is also examined to evaluate how good the developed forecasting model is. It is tested based on the value of the mean squared error (MSE), Mean Absolute Percent Error (MAPE) and Mean Absolute Deviation (MAD). The lower the value of error measure, the higher the accuracy of the model. The result shows that fuzzy time series model developed for primary school enrollments is the most accurate with the lowest error measure, with the MSE value being 0.38, MAPE 0.43 and MAD 0.43 respectively.

Keywords: Fuzzy time series, School Enrolments, Forecasting, Education Level, Accuracy.

INTRODUCTION

It is very crucial to make reasonably accurate forecasting of the future school enrolment in Malaysia because many decisions of the education system and resource planning will be made from them. For this reason, many researchers have proposed various methods to predict school enrolments. Fuzzy time series is one of the popular approaches used by researchers in forecasting problems due to its high accuracy and effectiveness compared to the other methods. It is an idea formulated by Song and Chissom proposed in 1993. It has many applications such as, forecasting wheat production (Stevenson & Porter, 2009), forecasting rainfall distribution of a region (Dani & Sharma, 2013), forecasting short-term electric load proposed by Huang (2015) and etc.



A higher order fuzzy time series forecasting model based on adaptive expectation and ANN was suggested by Hakan et al., (2010). The forecasted values have been adjusted through adaptive expectation and feedforward neural network in ANN. Fuzzy relationships in the higher order fuzzy time series are defined easily using this approach. As a result, the forecasting precision has increased significantly as presented in the results. Chen et al., (2009) have used fuzzy time series in projecting the enrolments of the University of Alabama. The automatic clustering techniques and fuzzy logical relationships were used to analyse the data. This method has a higher accuracy with a smaller MSE value and was found to be more accurate than the previous methods. Mahdzar et al., (2015) used fuzzy time series in predicting the number of tourist arrivals in Sabah. The study was conducted to assist the government in effectively planning strategies that would maintain and increase the production of tourism in Malaysia. P. Ramesh & K. Razak used a narrative time-invariant fuzzy time series for forecasting college admissions. They have found that fuzzy time series have significantly impact on achieving a better forecasting accuracy.

From the literature, we can see that researchers have used fuzzy time series in various fields. Therefore, this paper is carried out to show the application of fuzzy time series in projecting number of school enrolments in Malaysia.

METHODOLOGY

In this study, secondary data is taken from the United Nations Educational, Scientific, and Cultural Organization (UNESCO) institute. It consists of statistics of gross school enrolments in all stages of education in Malaysia from the year 1981 to 2018; a time-series data. The 35-years data from 1983 to 2018 is treated as the estimation period which is then used to forecast the number of school enrolments for the preceding years. The data is chosen just for simulation purposes to demonstrate the accuracy of the methods applied.

According to Hakan et al. (2014), fuzzy time series approach can deal with very small data and does not require the linearity assumption. This method makes the process of calculation become straightforward, (Dani & Sharma, 2013) and easy to apply in so many problems. There are seven steps to be followed. It can be summarized as follows:

Step 1: All data for pre-primary, primary, secondary, and tertiary schools for the year 1981 until 2018 is imported into MS Excel. There are 38 sets of data for each educational level, thus there are 152 sets of data in total. The data is analysed and converted into percentage using MS Excel based on equation 1.1.

$$y_n = \frac{(y_n - y_{n-1})}{y_{n-1}} \times 100, \quad (1.1)$$

where, y_n = total number of school enrolment in the current year, and y_{n-1} = total number of school enrolment in the previous year. For example, the change of school enrolment from 1981 to 1982 was calculated according to equation (1.1) as follows:

$$Change_{s_{1982}} = \frac{30.55 - 21.73}{30.55} \times 100 = 40.59$$

Step 2: Based on the calculation in Step 1, the universe of discourse, $U = [D_{min} - D_1, D_{max} + D_2]$ and dissolution U is defined into n with equal length of an interval as u_1, u_2, \dots, u_n . The value of D_{min} and



D_{max} are identified to define U . Next, the universe of discourse is defined as $[D_{min} - D_1, D_{max} + D_2]$ where D_1 and D_2 are two proper positive numbers to be assigned in U . The actual values are shown in Table 1 below.

Table 1: The minimum value, maximum value, the value of D_1 , and value of D_2 for all levels of education

Level of education	Minimum value (%)	Maximum value (%)	Value of D_1	Value of D_2
Pre-primary school	-24.57	49.46	5.43	0.54
Primary school	-1.72	2.94	0.28	0.06
Secondary school	-4.12	12.31	0.88	0.69
Tertiary school	-10.57	38.81	0.43	1.19

Based on the values in Table 1, the universe of discourse, $U = [D_{min} - D_1, D_{max} + D_2]$, was determined in this step as shown in Table 2.

Table 2: Table of universe of discourse

Level of education	Universe of discourse
Pre-primary school	$U = [-30, 50]$
Primary school	$U = [-2, 3]$
Secondary school	$U = [-5, 13]$
Tertiary school	$U = [-11, 40]$

Step 3: The fuzzy set U_i is partitioned into several length intervals. In this study, i can be anywhere in the range between 1 and 7 because the data in each education level are different every year. Then, the fuzzification of interval and the frequency distribution of each interval is calculated. The length of the interval for fuzzification is calculated using Equation (1.2).

$$\text{Length of interval} = \frac{(D_{max} + D_2) + (D_{min} - D_1)}{7}, \quad (1.2)$$

The actual calculated values are as follows:

$$\text{Length of interval for pre-primary school} = \frac{50 - (-30)}{7} = 11.43$$

$$\text{Length of interval for primary school} = \frac{3 - (-2)}{7} = 0.71$$

$$\text{Length of interval for secondary school} = \frac{13 - (-5)}{7} = 2.57$$

$$\text{Length of interval for tertiary school} = \frac{40 - (-11)}{7} = 7.29$$

After that, the intervals are fuzzified, where the value of this length is added into each interval. The actual values of interval length for all education stage are presented in Table 3.

Table 3: The length of interval for pre-primary and primary school

Pre-primary school	Primary school	Secondary school	Tertiary school



$U_1 = [-30.00, -18.57]$	$U_1 = [-2.00, -1.28]$	$U_1 = [-5.00, -2.43]$	$U_1 = [-11.00, -3.71]$
$U_2 = [-18.57, -7.14]$	$U_2 = [-1.28, -0.57]$	$U_2 = [-2.43, 0.14]$	$U_2 = [-3.71, 3.57]$
$U_3 = [-7.14, 4.29]$	$U_3 = [-0.57, 0.15]$	$U_3 = [0.14, 2.71]$	$U_3 = [3.57, 10.86]$
$U_4 = [4.29, 15.72]$	$U_4 = [0.15, 0.86]$	$U_4 = [2.71, 5.29]$	$U_4 = [10.86, 18.14]$
$U_5 = [15.72, 27.15]$	$U_5 = [0.86, 1.57]$	$U_5 = [5.29, 7.86]$	$U_5 = [18.14, 25.43]$
$U_6 = [27.15, 38.57]$	$U_6 = [1.57, 2.29]$	$U_6 = [7.86, 10.43]$	$U_6 = [25.43, 32.71]$
$U_7 = [38.57, 50.00]$	$U_7 = [2.29, 3.00]$	$U_7 = [10.43, 13.00]$	$U_7 = [32.71, 40.00]$

After the fuzzification process, the frequency distribution is generated for each interval as shown in Table 4.

Table 4: Frequency distribution of each interval for pre-primary and primary school

Intervals	Numbers of data			
	Pre-primary school	Primary school	Secondary school	Tertiary school
U_1	1	4	5	3
U_2	3	2	12	10
U_3	18	12	12	13
U_4	11	10	3	7
U_5	1	5	3	2
U_6	1	2	1	1
U_7	2	2	1	1

The equal length of the interval is divided by the numbers of data for each interval. An example of calculation is shown below:

v_i for Pre-primary school:

$v_1 = [-30.00, 18.57)$ with the length of 11.43,

$v_2 = [-18.57, -14.76), \dots, v_4 = [-10.95, -7.14)$ with the length 3.81, etc.

Step 4: The number of intervals will be following the number of linguistic variables which, in this study, used trapezoidal fuzzy sets A_1, A_2, \dots, A_m as represented below.

$$A_1 = [b_0, b_1, b_2, b_3],$$

$$A_2 = [b_1, b_2, b_3, b_4],$$

-

$$A_n = [b_{n-1}, b_n, b_{n+1}, b_{n+2}].$$

An example of actual interval generated in fuzzy trapezoidal for pre-primary school is as follows:

$$A_1 = [-33.81, -30.00, -18.57, -14.76],$$

$$A_2 = [-30.00, -18.57, -14.76, -10.95],$$

$$A_3 = [-18.57, -14.76, -10.95, -7.14],$$

-



$$A_{36} = [27.15,38.57,44.29,50.00],$$

$$A_{37} = [38.57,44.29,50.00,55.72].$$

Step 5: All the data is listed in percentage and being classified based on the interval that has been generated in Step 4. The fuzzy set A_i shows a linguistics value and if the data is founded in the range of v_j , then it will be a fuzzy number, A_j . Then, the fuzzy logical relationships are generated based on the classified data. Fuzzy logical relation is symbolised as follows: $A_i \rightarrow A_j$, where A_i is the present state and A_j is the upcoming state. The actual fuzzified data is produced and categorized into the corresponding fuzzy number as shown in Table 5.

Table 5: An example of fuzzified total school enrolment for pre-primary school

Year	Total enrolment	Changes (%)	Fuzzified total enrolment
1981	21.73		
1982	30.55	40.59	A37
1983	32.64	6.84	A26
1984	33.12	1.47	A19
.	.	.	.
.	.	.	.
2016	96.94	-0.08	A17
2017	99.97	3.13	A22
2018	99.22	-0.75	A16

Then, fuzzy logical relationship is also constructed for all datasets. An example of logical relationship is shown in Table 6.

Table 6: Fuzzy logical relationship for pre-primary school

$A_{37} \rightarrow A_{26}$	$A_5 \rightarrow A_{28}$	$A_{37} \rightarrow A_2$	$A_{10} \rightarrow A_{18}$	$A_5 \rightarrow A_{28}$	$A_{23} \rightarrow A_{25}$
$A_{26} \rightarrow A_{19}$	$A_{28} \rightarrow A_{13}$	$A_2 \rightarrow A_4$	$A_{18} \rightarrow A_{18}$	$A_{28} \rightarrow A_{35}$	$A_{25} \rightarrow A_{23}$
$A_{19} \rightarrow A_{25}$	$A_{13} \rightarrow A_{30}$	$A_2 \rightarrow A_4$	$A_{18} \rightarrow A_{25}$	$A_{35} \rightarrow A_{22}$	$A_{23} \rightarrow A_{24}$
$A_{25} \rightarrow A_{12}$	$A_{30} \rightarrow A_{19}$	$A_4 \rightarrow A_{24}$	$A_{25} \rightarrow A_{27}$	$A_{22} \rightarrow A_{21}$	$A_{24} \rightarrow A_{17}$
$A_{12} \rightarrow A_{24}$	$A_{19} \rightarrow A_{20}$	$A_{36} \rightarrow A_{11}$	$A_{27} \rightarrow A_{22}$	$A_{21} \rightarrow A_{21}$	$A_{17} \rightarrow A_{22}$
$A_{24} \rightarrow A_5$	$A_{20} \rightarrow A_{37}$	$A_{11} \rightarrow A_{10}$	$A_{22} \rightarrow A_5$	$A_{21} \rightarrow A_{23}$	$A_{22} \rightarrow A_{16}$

Step 6: Fuzzy relationship groups are generated in this step. The fuzzy logical relationships' rule is also arranged in groups as follows:

Table 7: Some of Fuzzy logical relationship based on their rule for pre-primary school

Fuzzy set	Relationship	Rule
$A_{37} \rightarrow A_{26}, A_2$	One to many	Rule 3



$A_{26} \rightarrow A_{19}$	One to one	Rule 2
$A_{19} \rightarrow A_{25}, A_{20}$	One to many	Rule 3
$A_{25} \rightarrow A_{12}, A_{27}, A_{23}$	One to many	Rule 3
.	.	.
.	.	.
$A_{23} \rightarrow A_{25}, A_{24}$	One to many	Rule 3
$A_{17} \rightarrow A_{22}$	One to one	Rule 2

Step 7: Each fuzzy relationship rule group is classified into one of the three different types of rule. The predicted production for each group is different according to the rule set. The rules involved in this step are shown as follows:

Rule 1: The fuzzy group of A_j is empty, which means A_j has no relationship with others. It can be represented as $A_j \rightarrow \emptyset$ or $A_j \rightarrow A_j$. The forecast value for this rule is $F_{vr} = R[NSTFN(A_j)]$

Rule 2: The fuzzy group of A_j is one to one, which means there is only one relationship with A_j and can be written as $A_j \rightarrow A_m$. The forecast value is $F_{vr} = R[NSTFN(A_m)]$.

Rule 3: The fuzzy group A_j is one to many. For example:

$A_j \rightarrow A_{m1}, A_j \rightarrow A_{m2}$ and $A_j \rightarrow A_{m3}$. It is written as $A_j \rightarrow A_{m1}, A_{m2}, A_{m3}$, and the forecast value will be calculated as follows:

$$NSTFN(A_j) = \left[\begin{array}{l} t_2 + \left[\frac{t_4 - t_1}{4} \right] - \frac{t_4 + t_1}{2}, t_2 + \left[\frac{t_4 - t_1}{4} \right], \\ t_3 + \left[\frac{t_4 - t_1}{4} \right], t_3 + \left[\frac{t_4 - t_1}{4} \right] + \frac{t_4 + t_1}{2} \end{array} \right] \quad (1.3)$$

$$F_{vr} = R \left[\frac{NSTFN(A_{m1}) + NSTFN(A_{m2}) + NSTFN(A_{m3})}{n} \right] \quad (1.4)$$

where n is the number of A_i in this group.

Example of Calculation for Step 7

The data in this study are classified into two type of rules, which are Rule 2 and Rule 3. For Rule 2, the fuzzy logical relationship group for the year 1995 to 1996 is $A_2 \rightarrow A_4$. Step 4 was repeated to get the trapezoidal number of A_4 using the following calculation:

$$A_4 = [-14.76, -10.95, -7.14, -6.51]$$

$$t_2 = -10.95$$

$$t_3 = -7.14$$

$$t_1 = -10.95 - (-14.76)$$

$$t_4 = -6.51 - (-7.14)$$

$$= 3.81$$

$$= 0.63$$



$$\begin{aligned} NSTFN(A_4) &= \left[t_2 + \left[\frac{t_4 - t_1}{4} \right] - \frac{t_4 + t_1}{2}, t_2 + \left[\frac{t_4 - t_1}{4} \right], \right. \\ &\quad \left. t_3 + \left[\frac{t_4 - t_1}{4} \right], t_3 + \left[\frac{t_4 - t_1}{4} \right] + \frac{t_4 + t_1}{2} \right] \\ &= [-13.97, -11.75, -7.94, -5.72] \end{aligned}$$

$$\begin{aligned} F_{vt} &= \frac{-13.97 + (-11.75) + (-7.94) + (-5.72)}{4} \\ &= -18.57\% \end{aligned}$$

Based on F_{vt} which was calculated for the forecast value for the year 1996, the fuzzified total school enrolment for Rule 3 is:

$$\begin{aligned} A_{19} &\rightarrow A_{25}, A_{20} & A_{25} &= [5.33, 6.37, 7.4, 8.44] \\ t_2 &= 6.37 & t_3 &= 7.4 \\ t_1 &= 6.37 - 5.33 & t_4 &= 8.44 - 7.4 \\ &= 1.04 & &= 1.04 \\ A_{20} &= [1.75, 2.38, 3.02, 3.65] \end{aligned}$$

$$\begin{aligned} t_2 &= 2.38 & t_3 &= 3.02 \\ t_1 &= 2.38 - 1.75 & t_4 &= 3.65 - 3.02 \\ &= 0.63 & &= 0.63 \end{aligned}$$

$$\begin{aligned} NSTFN(A_{25}) &= \left[t_2 + \left[\frac{t_4 - t_1}{4} \right] - \frac{t_4 + t_1}{2}, t_2 + \left[\frac{t_4 - t_1}{4} \right], \right. \\ &\quad \left. t_3 + \left[\frac{t_4 - t_1}{4} \right], t_3 + \left[\frac{t_4 - t_1}{4} \right] + \frac{t_4 + t_1}{2} \right] \\ &= [5.33, 6.37, 7.40, 8.44] \end{aligned}$$

$$\begin{aligned} NSTFN(A_{20}) &= \left[t_2 + \left[\frac{t_4 - t_1}{4} \right] - \frac{t_4 + t_1}{2}, t_2 + \left[\frac{t_4 - t_1}{4} \right], \right. \\ &\quad \left. t_3 + \left[\frac{t_4 - t_1}{4} \right], t_3 + \left[\frac{t_4 - t_1}{4} \right] + \frac{t_4 + t_1}{2} \right] \\ &= [1.75, 2.38, 3.02, 3.65] \end{aligned}$$

Then,

$$\begin{aligned} F_{vt} &= R \left[\frac{NSTFN(A_{m1}) + NSTFN(A_{m2}) + NSTFN(A_{m3})}{n} \right] \\ F_{vt} &= R \left[\frac{5.33 + 1.75, 6.37 + 2.38, +7.40 + 3.20, 8.44 + 3.65}{2} \right] \\ &= [3.54, 4.38, 5.21, 6.05] \end{aligned}$$



$$F_{vt} = R \left[\frac{3.54 + 4.38 + 5.21 + 6.05}{4} \right]$$

$$= 4.79\%$$

The process is repeated to forecast the number school enrolments of the remaining fuzzy logical relationships groups.

Accuracy Test

In this study, the accuracy of the forecasted value is evaluated using Mean Square Error (MSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Deviation (MAD). The value of MSE is calculated using equation (1.5), MAPE using equation (1.6) and MAD using equation (1.7).

$$e_t = y_t - \hat{y}_t,$$

$$MSE = \frac{\sum_{t=1}^n e_t^2}{n}, \quad (1.5) \quad MAPE = \frac{\sum_{t=1}^n |(e_t / y_t) \times 100|}{n} \quad (1.6) \quad MAD = \frac{\sum_{t=1}^n |e_t|}{n} \quad (1.7)$$

where, y_t = the actual value of total school enrolment at time t , \hat{y}_t = the forecasted value of total school enrolment and n = numbers of data. The values of MAPE, MSE and MAD are analyzed based on the lowest value of error measure.

FINDINGS AND DISCUSSIONS

The forecasted values in percentage are changed into the total value of the school enrolments. Table 8 shows the comparison between the actual number of school enrolments and the forecasted values for each year using FTS for all levels of schools in Malaysia. The forecasted value for the year 1981 and 1982 cannot be evaluated because there is no previous input data for these two years.

Table 8: The forecasted value for Pre-primary, Primary, Secondary and Tertiary school

Years	Pre-primary School		Primary School		Secondary School		Tertiary School	
	Total enrolment	Forecasted value	Total enrolment	Forecasted value	Total enrolment	Forecasted value	Total enrolment	Forecasted value
1981	21.73		96.34		57.19		3.82	
1982	30.55		95.91		57.33		3.99	
1983	32.64	28.92	95.51	95.85	60.79	62.14	4.33	4.27
1984	33.12	33.31	95.27	95.45	62.97	65.83	4.32	4.34
1985	35.19	34.71	95.64	96.07	63.18	63.06	4.81	4.86
1986	34.12	36.54	96.33	96.45	69.25	68.48	5.72	5.23
1987	35.64	36.12	95.9	95.98	68.06	68.46	6.73	7.35
1988	32.95	39.58	96.11	95.84	66.33	67.47	7.35	7.40
1989	35.79	36.25	94.69	94.83	65.39	65.85	7.38	7.41
1990	34.86	41.10	94.79	95.23	63.41	63.58	6.6	6.84
1991	38.86	39.07	94.61	94.94	64.53	64.65	7.18	7.22
1992	39.42	39.66	94.4	94.54	65.61	64.69	7.88	7.95
1993	40.11	41.31	95.37	95.19	64.53	65.77	8.91	9.54
1994	59.95	59.02	96.53	96.40	64.78	63.97	9.22	8.96
1995	45.22	56.76	97.15	97.12	64.82	66.43	10.14	10.21
1996	40.02	40.77	97.25	97.33	66.43	66.61	12.97	12.28
1997	41.95	42.36	97.48	97.40	70.01	70.80	15.59	14.88
1998	56.54	46.59	97.82	97.16	78.63	75.81	21.64	20.03
1999	54.5	54.84	96.14	96.30	77.85	77.73	22.62	24.83
2000	52.05	52.51	95.86	96.01	77.41	77.54	25.62	24.19



2001	52.38	52.79	98.68	96.98	76.9	77.44	25	25.77
2002	52.67	54.56	98.08	98.54	77.02	76.92	27.44	26.89
2003	55.68	54.86	96.6	97.49	82.9	83.48	30.61	33.23
2004	59.86	61.84	97.37	97.15	84.13	84.35	30.03	33.27
2005	61.89	62.30	99.72	100.12	80.66	83.07	27.92	28.42
2006	57.19	61.04	99.99	99.58	79.73	79.82	28.41	28.80
2007	62.04	62.91	98.69	99.66	77.48	79.47	30.01	30.14
2008	74.40	71.25	97.95	98.08	77.11	76.93	33.44	33.74
2009	76.91	77.43	97.98	97.36	76.12	76.72	35.49	36.35
2010	79.14	75.85	98.66	98.72	77.27	75.87	37.03	37.25
2011	81.23	82.40	99.76	99.86	77.47	76.30	36.15	39.59
2012	84.67	84.58	101.08	100.84	81.48	83.97	37.61	38.88
2013	89.41	90.06	102.74	103.21	81.87	82.03	39.07	38.74
2014	93.01	92.83	104.66	103.75	85.35	83.96	39.51	40.24
2015	97.02	98.93	104.36	105.69	84.93	85.47	45.59	46.06
2016	96.94	107.75	103.82	105.58	84.71	84.50	46.76	47.03
2017	99.97	100.89	105	105.08	85.45	85.65	43.72	44.26
2018	99.22	98.60	105.29	106.13	81.99	82.84	45.13	45.09

Figure 1 illustrates the ups-and-down movement of the data between the actual and the forecasted values. The horizontal axis of the graph displays the timeframe from the year 1981 to 2018 and the vertical axis represents the total numbers of school enrollments. In overall movement, it shows that most of the forecasted values are closest to the actual data. Hence, to be more specific, the error values are calculated.

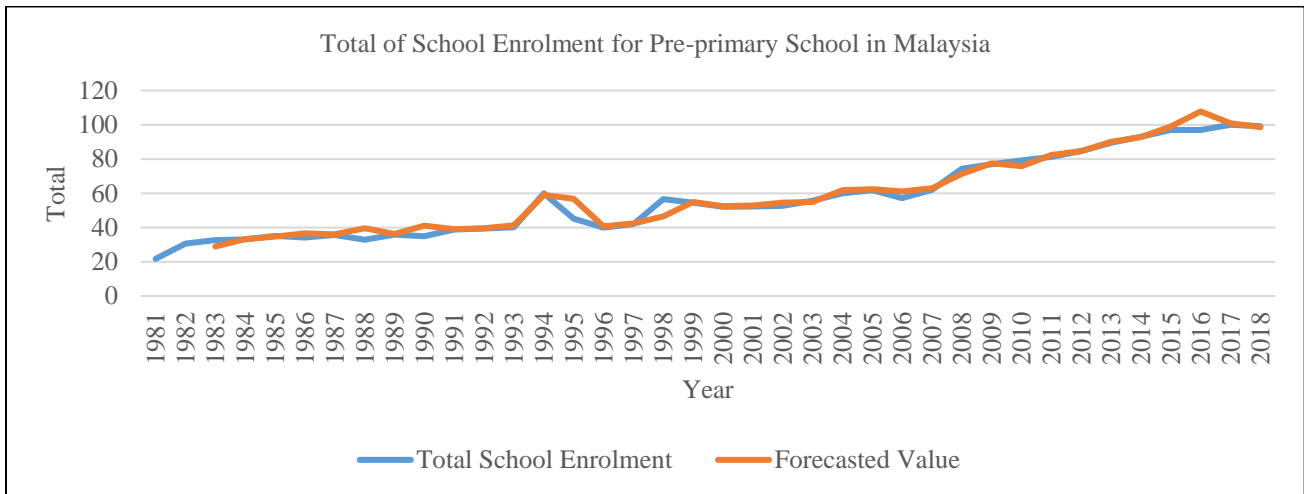


Figure 1: The graph of the actual and forecasted value for pre-primary school

Table 9 shows the values of MSE, MAPE and MAD for four different level of education.

Table 9: The value of MSE, MAPE and MAD

Education Level	MSE	MAPE	MAD
Pre-primary	14.10	4.4	2.23



Primary	0.38	0.43	0.43
Secondary	1.39	1.19	1.68
Tertiary	1.28	3.35	0.72

The model for primary school enrolment has produced the lowest MSE, MAPE and MAD which are 0.38, 0.43 and 0.43 respectively. This indicates that the forecasted values of fuzzy time series model are most accurate for the data of primary school enrolment. Meanwhile the highest error was produced for the data of Pre-primary school. This result might be affected by the situation where the enrolments of children for pre-primary schools are influenced by many unpredictable factors.

CONCLUSION AND RECOMMENDATIONS

The important part of the prediction is identifying the smallest error to determine the most accurate mathematical model that can fit the actual data. In this study, fuzzy time series approach is used to generate the mathematical model of school enrolment for pre-primary, primary, secondary, and tertiary school in Malaysia from 1981 to 2018. The actual data and the forecasted data are analyzed to determine the difference between these two data. By calculating the lowest value of MSE, MAPE and MAD for each stage of education, the mathematical model with the best accuracy was determined. The best result came from the model constructed for the data of primary school enrolment which has the lowest values of MSE, MAPE and MAD which are 0.38, 0.43 and 0.43 respectively. In overall, the values of error measures of model for all education level are still reasonable and acceptable, indicating that fuzzy time series approach is an effective and accurate way to predict the number of school's enrolments.

This study has been done with only one method which is fuzzy time series approach. In future it might be possible to analyze the same data by using other approaches such as Artificial Neural Network (ANN) or a hybrid method which is a combination of two different models, which might produce a better result.

REFERENCES

- Chen, S. M. (2002). Forecasting enrollments based on high-order fuzzy time series. *Cybernetics and Systems*. <https://doi.org/10.1080/019697202753306479>
- Dani, S., & Sharma, S. (2013). Forecasting rainfall of a region by using fuzzy time series. *Asian Journal of Mathematics and Applications*, 2013(1), 1–10.
- Hakan, C., Egrioglu, E., Yolcu, U., & Uslu, V. R. (2014). A high order seasonal fuzzy time series model and application to international tourism demand of Turkey, *Journal of Intelligent & Fuzzy Systems*, 26, 295–302. <https://doi.org/10.3233/IFS-120738> Chen et al.
- Hakan, C., Yolcu, U., & Egrioglu, E. (2010). A high order fuzzy time series forecasting model based on adaptive expectation and artificial neural networks. *Mathematics and Computers in Simulation*, 81(4), 875–882. <https://doi.org/10.1016/j.matcom.2010.09.011>
- Huang, Y. (2015). Weighted Fuzzy Time Series Model for Load Forecasting. *NCITPA*, 81–90.
- Mahdzar, M., Ling, S. M., Nair, M. B., & Shuib, A. (2015). Forecasting tourist arrivals to Sabah using fuzzy time series. *International Conference on Natural Resources, Tourism and Services Management 2015*.



Ramesh, P., Razak, K. (2018). Forecasting Enrollment using Fuzzy Time Series. *International Journal of Pure and Applied Mathematics*, 2018(15), 3309- 3313

Stevenson, M., & Porter, J. E. (2009). Fuzzy Time Series Forecasting Using Percentage Change as the Universe of Discourse. *World Academy of Science, Engineering and Technology*, 55, 154–157.

