

Forecasting Enrolment Data of Surigao del Sur State University, Philippines using Regression Analysis and Multiplicative Decomposition Model

Esmael V. Maliberan^{a,1,*}

^a CITE Department, Surigao del Sur State University, Tandag City, Surigao del Sur, 8300 Philippines

¹ malibs_28@yahoo.com

* corresponding author

ARTICLE INFO

Article history

Received February 10, 2021

Revised March 7, 2021

Accepted April 10, 2021

Keywords

multiplicative decomposition

forecasting

trend

regression

enrolment data

ABSTRACT

The study endeavored to predict the number of students who will enroll in Surigao del Sur State University using the historic data from 2010-2015. It was accomplished to utilize two existing methods to form its model in forecasting the enrolment data of the University. To project the number of students enrolled at the university, regression analysis, and a multiplicative decomposition model is used. The Mean Absolute Percentage Error (MAPE) of the forecasted enrolment data was 3.10 percent, according to the results. This only demonstrated that the estimated data were similar to the actual data, implying that it is accurate. The predicted enrolment data can be used to support decision-making and as an input to the University Development Plan.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Introduction

The administration of Surigao del Sur State University faces many challenges as to the hiring of faculty members, especially during the enrolment period. This is because there is no available mechanism or model being used to accurately forecast the number of students who will enroll every semester. Accuracy in forecasting plays a vital role in decision making that is why several studies focused on the accuracy level of the model to predict historical data. [1-5]. [6] Proposed a hybrid algorithm based on fuzzy time series and genetic algorithms to predict enrolment, the proposed algorithms have a strong forecast result with a higher accuracy rate. However, there are only relatively few studies on enrolment forecasting in the Philippines. This study plays a vital role for the school administrators for proper planning and preparation to cater to students' need, particularly during the enrolment period. This study focuses on probing the enrolment data of Surigao del Sur State University from preceding years and applies time series analysis to predict the number of students who will study in the university. [7] utilized Markov Weighted Fuzzy Time-Series Model to predict the pollution in air using pollution index (API). It was also accepted by the University of Alabama to benchmark enrolment details which showed its accuracy in predicting enrolment data. [8] utilized Regression Analysis to predict what are the different factors distressing the enrolment data of some tertiary schools. The models were used to examine the variables and their relationships. Males are 23 percent less likely than females to attend school, according to enrolment forecasts based on statistical models. [9] Time series models were utilized to assess the influx of students enrolled in a course as a predictive aid for the Electronic School Management System. Although the presence of School Management Systems is available in most of the schools and universities, however, its functionality is limited only to admission, academic transactions, curricula, pre-registration,

registration, records management, and grading. Most of these systems do not support enrolment forecasting so that the management can predict the whole students enrolled in the program. This is why, during the enrolment era, many students face problems such as course unavailability, combined courses, or dissolved parts. The data used in this study is the entire population of students enrolled in Surigao del Sur State University from 2010-2015. Two models will be used in forecasting enrolment. These are the Linear Regression and the multiplicative decomposition model. Through the use of these models, this study will calculate the moving average, slope, and intercept in Enrolment results.

1.1. Theoretical Framework

This chapter presents different theories and concepts that were used to describe the structure of this study. The researcher will discuss the algorithm and analysis used to show the preferred actions to be taken in the study.

1.2. Theory on Forecasting

Intrinsically, forecasting is intertwined with decision-making. By helping agents to make decisions about an unpredictable future, forecasts serve their function. To the degree that different forecasts influence different decisions, forecast errors create costs for the decision-maker. As forecast errors in a random world are inevitable, the classical forecasting theory builds on the premise that agents want to minimize the potential costs associated with these errors. [10].

1.3. The theory of decision making

If the prediction of the upcoming value of a mutable or the likelihood of an occurrence is used for preparation or decision-making purposes, the decision-priorities, maker's opportunities, and limitations will all be included in the later evaluation of a prediction and the later contrast of alternative prediction.

1.4. Time Series Analysis

This research is based on the Time Series model. A time series is a group of observations x_t that are all recorded at the same time t . A discrete time series is when the discrete set T_0 of times by which interpretations are prepared is the separate set. For example, when interpretations are prepared at fixed time intervals. When measurements are continuously collected over a certain time interval, continuous time series are obtained [11]. When there is a durable surge or reduction in the results, a pattern occurs. It need not be linear. We can often refer to a "changing path" trend when it may go from a growing trend to a declining trend. When a sequence is affected by seasonal causes, a seasonal trend occurs (e.g., the quarter of the year, the month, or the day of the week). Seasonality is often a time that is set and understood. When data display rises and falls that are not of fixed length, a cyclic pattern occurs. Usually, the length of these variations is at least two years.

1.5. Multiplicative Decomposition

The magnitude of both the seasonal and periodic variations increases in several time series as the degree of the pattern increases. A multiplicative model is typically suitable in this case. Time series analysis is represented in the multiplicative model as the product of trend components, seasonal and irregular [12].

Observed series = Trend x Seasonal x Irregular

or

$$O_t = T_t \times S_t \times I_t$$

The seasonally adjusted data then becomes:

Seasonally Adjusted series = Observed / Seasonal = Trend x Irregular

or

$$\begin{aligned} SA_t &= \frac{O_t}{\hat{S}_t} \\ &= T_t \times I_t \end{aligned}$$

1.6. Related Literature

The context of this study used regression analysis and a multiplicative decomposition model in forecasting enrolment through time. [6] To forecast enrolment based on genetic algorithms and fuzzy time series, a hybrid algorithm was suggested. A good forecast result with a higher precision rate is presented by the proposed algorithms. This research uses the University of Alabama's historical enrolment from 1948 to 2010 to explain the forecasting process.

According to [13], A fuzzy time series was developed for efficient enrollment forecasting. The model consists of four steps: the description of the discourse and interval universe, the fuzzification of past data, the design of fuzzy relationships, and the prediction for the enrolment. The max-min operator was used as a discourse universe, and we contrasted our proposed system with the current linear method. The historical enrolment figures of the University of Agriculture, Abeokuta, were used as a data set for study and were implemented using Visual Basic. The fuzzy time series approach probable result is matched to that of the current least square method, with the fuzzy time series approach creating the smallest mean square error (MSE) values as compared to the least square method. The request was also utilized to guess the enrolment data for five years ahead. The proposed method was found to produce more precise results for forecasting than the current method. [9] They created software and combined it with their current Electronic School Management System as part of their research on Enrolment Forecasting Using Time Analysis. It assists academic program administrators in estimating the number of students expected to enroll in a subject and deciding the subjects to be offered with the help of a statistical method. Models were developed using MSU-IIT data from 1998 to 2009 and tested using real data from the 2009-2010 school year. [8] has done more studies on enrollment forecasting. The factors affecting a tertiary school's enrolment were predicted using regression analysis. The enrollment forecast was created using three models: the logic model (for demographic profile), the percentage, base, and rate principles (for price), and the Markov analysis (for enrollment) (for quality and convenience). Using the models, variables, and relations among them were investigated. [14] The development of the weight fuzzy time series was provided based on a set of chronological number variations in the Fuzzy Logical Group (FLG). It aimed to establish an acceptable weight for forecasting trend series data on fuzzy time series. For forecasting, a university enrollment data collection was used for Alabama University and Universiti Teknologi Malaysia (UTM). Results from this research showed that there was a great deal of progress in the proposed method. The Mean Square Error (MSE) and average error were used for the forecast fitness function.

According to [15], If time series interpretations comprise ambiguity, fuzzy time model techniques are applied. Furthermore, the expectations required for predictable time series methods do not need these methods. Fuzzy time series methods naturally contain three phases, specifically, fuzzification, fuzzy relationship determination, and defuzzification. At this point, artificial intelligence is widely used because of its accuracy and good performance. To discover the best interval lengths and regulate the outcome of the mutation operator, the authors suggested an efficient genetic algorithm. When likened to those found by other methods, the outcomes of using their novel method to actual datasets showed greater predicting accuracy.

The research [16] conducted an out-of-sample forecasting competition between GM (1,1), GM(1,1) rolling, and exponential smoothing in Forecasting efficiency of grey projection for education expenditure and school enrollment. Person forecasts were generated using annual time-series data from the National Center for Education Statistics (NCES) from 1991 to 2004. For the test of equivalent forecast accuracy, accuracy measurements were calculated using the MAD, MAPE, MSE, and F-statistics criteria. In terms of forecasting education spending and student enrollment, the GM(1,1) rolling model outperformed the other two models.

A new forecasting model was developed[17] based on two computational approaches, time-variant fuzzy logical relationship groups and the technique of K-mean clustering presented for academic enrolment. The K means clustering algorithm was used to separate historical data into clusters and conform them to different duration intervals. Fuzzification was then applied to all the historical data of the University of Alabama enrolments, based on the new intervals, and the projected output was determined by the proposed method. In predicting the total population of students enrolled at the University of Alabama from the 1971s to the 1992s, the proposed approach demonstrated greater accuracy.

2. Material and Method

There are few manuscripts published articles in journals related to BDA implementation. For instance, Batarseh and Latif in 2016 and Alani and team in 2016 assessed the quality of service and healthcare organization using BDA in healthcare industries in the United States and Iraq [21]; Moyne and Iskandar in 2017 implemented BDA in a manufacturing company [22]; [23]; Naimur Rahman and his team in 2016 used BDA to predict total electricity forecast in the United States [24]; Honarvar and Sami in 2019 tried to find a suitable solution for urban development by using the opportunities of big data and present data related to urban computing with the aim of assessing the knowledge that can be obtained through integration of multiple independent data sources in Smart Cities [25]. Meanwhile, the position of this manuscript that makes it different from previous related works is that rarely were studies found in the literature covering BDA in a gold, silver, and precious metal industry. This manuscript also explores the customer behaviors and habits based on a large amount of data analysis in conducting the transaction in a developing country like Indonesia.

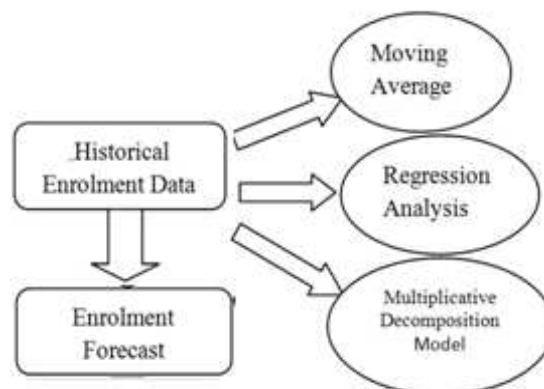


Fig. 1 Operational Framework of the Study

2.1. Forecasting Models

The university uses two semesters and a summer every year. This means that every year, there are three sets of enrolment data. By Gathering Enrolment data from 2010-2015, It's made up of data that has been collected over time. Since enrolment data follows a sensibly direct inclination and has a separate regular design of differences, a moving average is used as part of the pre-processing technique. This approach is applied to smooth enrolment data over time to get its trends. Two models are evaluated to forecast the enrolment of the university. The first method is to use simple linear regression to determine the slope and intercept. The seasonal, irregular, and pattern components of the enrolment data will be extracted using a Multiplicative Decomposition model.

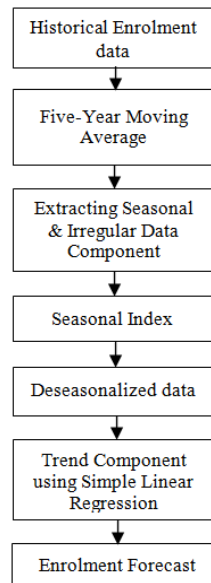


Fig. 2 System Architecture

Table 1. Result of Actual Historical Enrolment Data and Forecast Enrolment Data

year	AY	actual data	forecast data
1	1st sem	3500	3353
	2nd sem	2900	3081
	summer	700	774
2	1st sem	4001	4018
	2nd sem	3868	3655
	summer	904	910
3	1st sem	4955	4684
	2nd sem	4387	4228
	summer	1050	1046
4	1st sem	5299	5350
	2nd sem	4757	4802
	summer	1200	1182
5	1st sem	5798	6016
	2nd sem	5434	5376
	summer	1300	1317
6	1st sem		6682
	2nd sem		5950
	summer		1453

Based on Table I, there is a fluctuation in enrolment data for each semester. However, there is a consistent trend of increasing enrolment per semester per year. This means that the amplitude of both the seasonal (first semester) and irregular variations (second semester and summer) increase as the level of the trend rises. Figure 3 shows the graphical presentation of the actual past enrolment data and the predicted enrolment data.

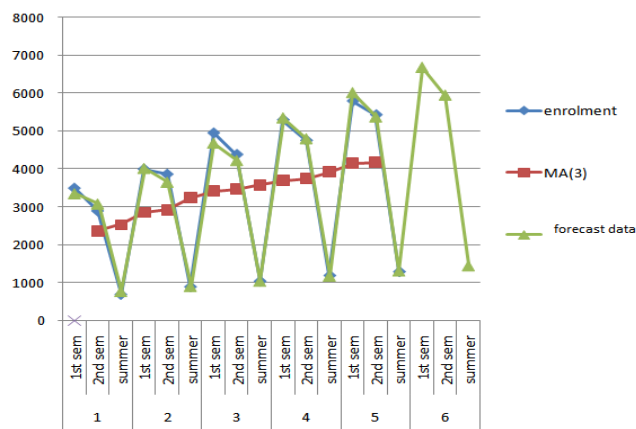


Fig. 3 Actual Enrolment Data vs. the Projected Enrolment Data of Surigao del Sur State University

2.2. Testing and Simulation

To assess the model's accuracy, differences or deviations between forecast and actual data are compared. Model evaluation is the last step in the model development process. The goal is to use mean absolute percentage error to compare and calculate the difference between real and forecasted values (MAPE). MAPE is a metric that assesses the accuracy of a trending time series value that has been fitted. MAPE is computed using the formula;

$$M = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

where

A_t - Actual value

F_t - Forecasted Value

The outcome of the testing is determined by comparing the number of students enrolled from 2010 to 2015 with the forecasted value for those years. Using MAPE, the estimated data from the time series models yields a 3.10 percent error. For example, the actual enrolment in the first semester of the academic year 2010-2011 is 3500, while the forecasted enrolment is 3353. MAPE is calculated by dividing projected data by actual data and subtracting projected data from actual data. This results in a 4.2 percent error for that semester. Obtaining the forecasting error is critical because it indicates the model's accuracy in the analysis. The researchers were able to forecast student enrolment for the academic year 2015-2016 based on a simulation using MSExcel. 6682 students will enroll in the first semester, 5950 students in the second semester, and 1453 students in the summer, according to the results.

2.3. Five Year Moving Average Method

The moving-average method is not only useful in leveling a time series in order to get its movement; it is the simple process that is used in calculating the periodic variation. Contrary to the least square technique, which articulates the movement in a mathematical equation ($Y' = a + bt$), the moving-average technique only flattens the variations in the data. The integer mean values are "moved" across the time series to do this.

The formula for the simple moving average of order 3

$$T_t = (Y_1 + Y_2 + Y_3) / 3$$

Where:

Y_1 - First Semester for Year 1

Y_2 - Second Semester for Year 1

Y_3 - Summer for Year 1

2.4. Regular and Irregular Component

The Regular and Irregular Variable is extracted by dividing the historical data of the enrollment into the moving average. And add up the first semester, second semester, and summer averages.

Table 2. Result of The Simulation of Actual Historical Enrolment Data and The Predicted Enrolment Data.

year	semesters	enrolment	MA(3)	seasonal component	deseasonalize	trend	forecast
1	1st sem	3500		1.54	2272.73	190.92	3353.00
	2nd sem	2900		1.06	2735.85	202.17	3081.00
	summer	700	2366.67	0.50	1400.00	213.41	774.00
2	1st sem	4001	2533.67	1.54	2598.05	224.65	4018.00
	2nd sem	3868	2856.33	1.06	3649.06	235.89	3655.00
	summer	904	2924.33	0.50	1808.00	247.13	910.00
3	1st sem	4955	3242.33	1.54	3217.53	258.37	4684.00
	2nd sem	4387	3415.33	1.06	4138.68	269.61	4228.00
	summer	1050	3464.00	0.50	2100.00	280.85	1046.00
4	1st sem	5299	3578.67	1.54	3440.91	292.10	5350.00
	2nd sem	4757	3702.00	1.06	4487.74	303.34	4802.00
	summer	1200	3752.00	0.50	2400.00	314.58	1182.00
5	1st sem	5798	3918.33	1.54	3764.94	325.82	6016.00
	2nd sem	5434	4144.00	1.06	5126.42	337.06	5376.00
	summer	1300	4177.33	0.50	2600.00	348.30	1317.00
6	1st sem			1.54	0.00	359.54	6682.00
	2nd sem			1.06	0.00	370.79	5950.00
	summer			0.50	0.00	382.03	1453.00

2.5. DE seasonalizing Data

Methods may be used to remove it from the time series to study the effect of other elements, such as cyclical and irregular variations, after determining the seasonal pattern. De-seasonalizing or seasonal change of data is the word used to describe removing the seasonal effect. The enrolment data is divided by the seasonal portion.

2.6. Importance of Regression

Linear regression is utilized to identify if a trend exists based on the attribute of the data.

Table 3. The Relationship Between Time (X) And Deseasonalized Data (Y)

t	deseasonalize
1	2302.63
2	2213.74
3	2258.06
4	2632.24
5	2952.67
6	2916.13
7	3259.87
8	3348.85
9	3387.10
10	3486.18
11	3631.30
12	3870.97
13	3814.47
14	4148.09
15	4193.55

To formally test whether a linear trend occurs, Table III data will be used. De seasonalized data as dependent variable (Y) and Time data as the independent variable (X). It is defined by the formula

Regression Formula:

$$Y = a + bX$$

where slope of trend line is calculated as:

$$b_1 = \frac{\sum (x - \bar{x}) * (y - \bar{y})}{\sum (x - \bar{x})^2}$$

and the intercept is computed as:

$$b_0 = \bar{y} - (b_1 * \bar{X})$$

Table 4. Result Of Simple Linear Regression Analysis

SUMMARY OUTPUT				
<i>Regression Statistics</i>				
Multiple R	0.9843764			
R Square	0.968997			
Adjusted R Square	0.9666121			
Standard Error	121.20764			
Observations	15			
<i>ANOVA</i>				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	1	5969272	5969272	406.3136
Residual	13	190986.8	14691.29	
Total	14	6160258		
<i>Coefficients</i>				
	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	
Intercept	2059.6464	65.85916	31.2735	1.27E-13
t	146.00968	7.243542	20.15722	3.44E-11

Table IV shows that the time coefficient is statistically significant; its p-value is = 3.44E⁻¹¹, which falls below 0.05. This indicates that there is sufficient evidence to support the existence of a data pattern. As a result, using a Simple Linear Regression model, historical SDSSU enrolment data can be forecasted.

Table 5. The Slope and Intercept

Intercept	2059.64
Slope	146.00

Since the slope and intercept will be used to find the pattern, they are very important values.

Trend

The formula can be used to measure trends once the intercept and slope have been determined.

Trend = Intercept + slope x time

Multiplicative Decomposition

Forecast enrolment data can be estimated using a multiplicative Decomposition model described by a formula once the pattern has been calculated.

$$O_t = T_t \times S_t \times I_t$$

Where O_t = Observed data/Forecast Data

S_t = Seasonal Component

I_t = Irregular Component

3. Conclusion

The study was able to approximate the number of students enrolled in university. It provided a graphical representation of both the actual and projected enrolment results. To forecast historical enrolment data, this study used a simple linear regression and multiplicative decomposition model. The accuracy calculation (MAPE) of 3.10 percent only showed that predicting enrolment data is similar to real data, and therefore the result has a higher degree of accuracy. The researcher assumes that this study should be implemented and incorporated into current school and university information systems. Other mathematical models may be incorporated into the current model to improve forecast accuracy.

References

- [1] SSG Abhishekn, and SR Singh, A score function-based method of forecasting using intuitionistic fuzzy time series, *New Math. Nat. Computer* 14 (1), 91-111, 2018.
- [2] Abhishekn, S. Gautam and SR. Singh, A refined method of forecasting based on high-order intuitionistic fuzzy time series data *Progress in Artificial Intelligence* vol.7, pp. 339-350, 2018.
- [3] SK. Bharati, Abhishekh, SR. Singh, A computational algorithm for the solution of fully fuzzy multi-objective linear programming problem, *International Journal of Dynamics and Control* vol. 6, no. 3, pp. 1384-1391, 2018.
- [4] S. Gautam and Abhishekn, A Novel Moving Average Forecasting Approach Using Fuzzy Time Series Data Set, *Journal of Control, Automation and Electrical Systems*, 2019.
- [5] R. Maliberan, Forecasting tourist arrival in the Province of Surigao del Sur, Philippines using Time Series Analysis, *International Journal on Informatics Visualization*, vol. 3, no. 3, 2019.
- [6] A. Talal, H. Ibraheem, and A. Salim, A Hybrid Algorithm to Forecast Enrolment Based on Genetic Algorithms and Fuzzy Time Series. *IAJIT First Online Publication*, 2013.
- [7] Y. Alyousifi, M. Othman, I. Faye, R. Sokkalingam and P. Silva, Markov Weighted Fuzzy Time-Series Model Based on an Optimum Partition Method for Forecasting Air Pollution, *International Journal of Fuzzy Systems* vol. 22, pp.1468–1486, 2020.
- [8] G. Dimaculangan, Using Regression Analysis to Forecast the Factors Affecting the Enrolment of a Tertiary School. *Asia Pacific Journal of Multidisciplinary Research*, vol. 2, no. 1, pp.121-130, 2014.
- [9] R. Lavilles, and M. Arcilla, Enrolment Forecasting for School Management System. *IJMO*, pp. 563-566, 2012.
- [10] S. Manganelli, A New Theory Of Forecasting. *European Central Bank Working Paper Series*, (584), 2006.
- [11] P. Brockwell, and R. Davis, *Introduction to Time Series and Forecasting*.
- [12] R. Hyndman, and G. Athanasopoulos, *Forecasting Principles and Practice*, 2018.
- [13] A. Ikuomola, and A. Adebisi, A comparison of fuzzy time series and least-square method in forecasting students' enrolment. *Journal of Natural Sciences, Engineering And Technology*, vol. 10, no. 1, pp. 19-30, 2011.
- [14] Z. Ismail, and R. Efendi, Enrollment Forecasting based on Modified Weight Fuzzy Time Series. *Journal Of Artificial Intelligence*, vol. 4, no. 1, pp. 110-118, 2011.
- [15] E. Bas, V. Uslu, U. Yolcu and E. Egrioglu, A Modified Genetic Algorithm for forecasting fuzzy time series, *Applied Intelligence*, vol. 41, no. 2, pp. 453-463, 2014.
- [16] H. Tang, and M. Yin, Forecasting performance of grey prediction for education expenditure and school enrollment. *Economics of Education Review*, vol. 31, no. 4, pp. 452-462, 2012.
- [17] N. Tinh, V. Vu and T. Linh, A New Method for Forecasting Enrolments Combining Time-Variant Fuzzy Logical Relationship Groups And K-Means Clustering. *International Research Journal of Engineering and Technology (IRJET)*, 03(03), 2016.