The Design of Polynomial Curve Fitting Model: A Prediction of Freshmen Enrolment and a Recommender System

Asia Pacific Journal of Multidisciplinary Research
Vol. 6 No.4, 38-44
November 2018 Part II
P-ISSN 2350-7756
E-ISSN 2350-8442
www.apjmr.com
CHED Recognized Journal
ASEAN Citation Index

Joan P. Lazaro[1], Melvin A. Ballera[2], Dennis B. Gonzales[3]
[1]Faculty Member, University of the East, Caloocan, [2]Assistant Director for Research, Technological Institute of the Philippines, Manila, [3]Graduate School Coordinator, University of the East, Manila, Philippines
lazaro.joan@yahoo.com[1], maballera@yahoo.com.ph[2], dbgonzales33@gmail.com[3]

Date Received: March 4, 2018; Date Revised: November 10, 2018

Abstract – *Nowadays, as far as mathematical knowledge is concerned, primary forecasting techniques are needed to aid institutions or companies to generate a strategy for the future. Forecasting is one of the important tools for developing strategic choices that is thought as a vital part for valuable organization. The main aim of this study is twofold: to design a Polynomial Curve Fitting-Model based on freshmen enrolment report and to develop a recommender system that will aid the Higher Education Institution management for decision making. The researcher used descriptive method of research using historical and documentary analysis, backed by respondents composed of school admission and registrar staff and some college officials and administrators. The researcher used participant observation and informal interview. The study shows that Polynomial Curve Fitting Model can also be used to forecast enrolment data and the factors that affect student enrolment but comparing it with the other time series forecasting techniques, its Mean Absolute Percentage Error (MAPE) is high. The results of this study will be used by the decision maker to achieve some sets of institutional goals and can be a basis for long-range planning. Therefore, it is important to note that enrolment forecasting is an iterative and collaborative process, with models being annually reviewed based on feedback from various constituents within the community.*

Keywords – *Data Mining, Polynomial Curve Fitting, Prediction, Recommender System*

INTRODUCTION

Students' enrolment projection offers significant figures for organization in planning and decision making for the Higher Education Institutions (HEI). Forecasting can assist the management to make supplementary models based on the historical figures gathered. Enrolment prediction is associated to examining the existing styles, comprehending the substantial effect on the admission and income results. This evidence is usable to the institutions to impact the upcoming tactics and asset resolutions [1]. Some of the main reasons for management to consider forecasting in their institutions are to understand the relationship between the historical, existing and data relationship between them, in which forecasting measures can provide a substantial level of accurateness with a lesser amount of cost, the amount of the past figures in which the significance of the evidence that is gauged for dependable resolution and lastly, the period to develop the projection.

The influences in the students' selection to enroll in a certain HEI are significant since an HEI can comprehend the purpose of the student in picking a specific HEI over another. The evidence attained can be utilized by academies to support in the improvement of an advertising strategy [2]. The image indicators of an academia can be used to serve as the basis for enticing potential applicants in a university [3].

There are two certain categories for forecasting methods, one is extrapolation which is used to stretch the current trend and the second is normative which is used past data for projection of an anticipated future [4]. The enrolment forecasting techniques is very significant for an HEI to verify brief and extensive period forecasts for assets and facilities development. Precise projections remain significant in times of change. The increase and decrease of the student does not just affect decision making techniques, but can provide organization observed instructional quality [5].

Polynomials are basically a function in mathematics that is mostly used in solving problems and developing models in the field of engineering and science. In numerous instances of solving certain problems an

equation in the polynomial form is developed. One process to be able to model data is curve fitting, this creates a model that has the smallest possible error even without passing through any of the provided points. Polynomial equations, exponential equations and power functions are common equations used for curve fitting. Simple and moving averages, exponential smoothing and exponential models and polynomial analysis are some of the common techniques used within curve fitting. Polynomial curve fitting is a mathematical method used by creating a best fit projection based on a series of data provided, with possible constraints. The process of approximating values among data item is known as interpolation, its simplest form is drawing a straight line from point A to point B, while a more complicated interpolation, data from supplementary items is utilized.

There are numerous methods utilized by HEI to estimate enrolment. Four categories of forecasting techniques are investigated. First, is the quantitative technique that is centred on chronological information. Second, is the contribution model that integrates the chronological data and rely on a connection between enrolments with other factors, or by means of methods that incorporate subjective finding instead of using measureable procedures. Third, is the alteration or recommendations for modifications or appraisals to previously established forecasting techniques. Lastly, is the qualitative technique by means of surveys to ascertain possible students' enrolment [6]. Though the selection of suitable forecast methods is reliant on the accessibility of the HEI records. The methods concerned in each forecasting will provide contribution on the long and short duration planning of the HEI.

Freshmen student enrolment prediction provides significant figures for making decisions and administration planning in an HEI. By this research, the administrators will come up with powerful strategic and tactical decision making. The results of this study can be used by the decision maker to achieve some sets of institutional goals and can be a basis for long-range planning.

The development of recommender system in academic institutions helps the administrator identify more effectively the process of attracting more students. It resulted to a powerful strategic and tactical decision making.

In this research, software was created with a mathematical means that support the educational program officers in projecting the quantity of expected enrollees in college. Models are based on the data from University of the East (UE) Manila and Caloocan during the school year S.Y. 2010 to 2015 and were assessed by means of the actual figures of S.Y. 2015-2016.

The researcher, as a faculty member of the Department of Computer Engineering at UE for so many years, is earnest in his advocacy to develop an enrolment projection model using Polynomial Curve Fitting technique that will help the administrators to have a deeper understanding of figures of enrolment for organization in terms of planning and making decisions. By this study they will come up with powerful strategic and tactical decision making. In addition, this study serves the students to have a better education system.

OBJECTIVES OF THE STUDY

This study is three-fold in its objectives: to determine the predictors that will directly affect the enrollment status of the freshmen students; to design the polynomial curve fitting-model based on enrollment report; to analyze how efficient is the forecasting model as basis to develop the recommender system.

MATERIALS AND METHODS

The researcher used the descriptive method of research using historical and documentary analysis, backed by respondents composed of school registrar and admission staff and some college officials and administrators. The researcher used participant observation and informal interview. The historical method was used as auxiliary tools to dig and describe the student's records upon enrolment.

The researcher acquired information mainly from University of the East Manila and Caloocan. Data gathered from the respondent includes the number of populations of student from school year 2010 to 2016. Another important data are the student profiles upon entry in the school like age, house hold monthly income, high school grade, type of school, National Career Assessment Examination (NCAE) result and entrance exam results. Finally, the researcher asked the result of survey showing the reasons why the student enrolled in college.

The researcher used the two main methods for the student profile (i) information gain attribute evaluation using ranking filter; and (ii) logistic regression; and (iii) mean for the factors influencing student enrolment.

The polynomial curve fitting model was designed through the historical data gathered by the researcher to predict the enrolment for the coming years. To effectively evaluate the polynomial curve fitting model, a comparison was done with the other time series

forecasting technique like the Simple Moving Average and Exponential Smoothing.

The recommender system was developed based on the results of the polynomial curve fitting model. This was used by the decision maker to achieve some sets of institutional goals which can be a basis for a long-range planning.

RESULTS AND DISCUSSION

A. Predictors that will directly affect the enrolment status of freshmen students

The researcher used the two main methods for the student profile (i) information gain attribute evaluation using ranking filter; and (ii) logistic regression; and (iii) mean for the factors influencing student enrolment.

1. Information Gain Attribute Evaluation using Ranking Filter

Information Gain Attribute Evaluation Method evaluates the worth of an attribute by measuring the information gain with respect to the class. The information gain of an attribute tells you how much information with respect to the classification target the attribute gives you. Based on the empirical testing, the ranking of attributes are shown in Table 1.

Table 1. Ranked Attributes Results using Information Gain Attribute Evaluation Method

Attributes	Ranked Attributes	Attributes Percentage
Entrance Results	1	0.17814
NCAE	2	0.06279
Monthly Income	3	0.04409
High School Grade (HSGrade)	4	0.04390
Age	5	0.03772
Working Parents	6	0.02447
High School Type (HSType)	7	0.02082
Parents Work Abroad	8	0.01698
High School Scholar	9	0.00400

The ranked attributes in Table 1 revealed that entrance exam results (17.814%) contribute a higher percentage of student's enrolment. Second in the rank is the score of students in National Career Assessment Examination (NCAE) with 6.279%. The monthly income of parents ranked third (4.409%) and the high school grade was ranked fourth (4.39%). This implies that the student will enroll if they passed the entrance examination.

2. Logistic Regression

Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous or binary. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. A logistic regression produces a logistic curve as shown in Figure 1, which is limited to values between 0 and 1.

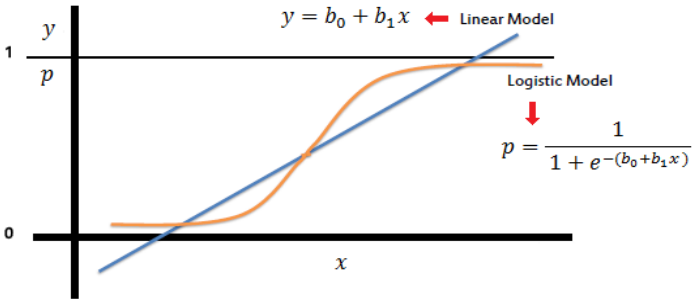

Figure 1. Logistic Regression Model

The logistic regression was used to determine the significant predictors. Recoding was done on the enrolment status variable such that 0=not enrolled and 1=enrolled. Recoding is essential for simple analysis of the result after implementing logistic regression. The recoding was done using SPSS v.21

Logistic regression analysis on all the hypothesized predictors was conducted. A logistic regression model where all predictors are significant ($p<.05$) was considered.

A total of nine (9) possible predictors were tested using logistic regression, which were used as valid and important predictors to attributes used for data mining purposes.

To determine the statistical significance of a predictor the p value was used. The predictor is statistically significant when a p-value is less than the significance level. The *p*-value is the probability of observing an effect given that the null hypothesis is true whereas the significance or alpha (α) level is the probability of rejecting the null hypothesis given that it is true. In practice significance level is chosen before data collection and is usually set to 0.05.

The model occurs in step 6. SPSS v.21 was utilized to do the calculation. A result of the binary logistic regression using forward method is summarized in Table 2.

Analysis of the data reveals that six variables significantly contributes to student enrolment namely Income (B=0.962, $p<0.05$), Age (B=-0.660, $p<0.05$),

High School Grade (B=-0.160, p<0.05), High School Type (B=-1.221, p<0.05), NCAE (B=-0.490, p<0.05) and Entrance exam (B=0.876, p<0.05).

Table 2. Variables in the Equation of Logistic Regression

Variables in the Equation		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 6[f]	Income	0.962	0.204	22.140	1	0.000	2.616
	Age	-0.660	0.121	29.803	1	0.000	0.517
	HSGrade	-0.160	0.059	7.263	1	0.007	0.852
	HSType	-1.221	0.397	9.455	1	0.002	0.295
	NCAE	-0.490	0.065	56.305	1	0.000	0.612
	Entrance Resutls	0.876	0.080	120.168	1	0.000	2.402
	Constant	-3.643	3.624	1.011	1	0.315	0.026

The B coefficients for income and entrance examination are positive, indicating that the higher the income of the parents and scores in the entrance examination the higher the odds of being enrolled.

Mean

Table 3 shows the comparison of the result of survey conducted by the Admission Office of UE Caloocan and Manila from school year 2010 to 2016. Using the mean, it shows that the significant predictors that will affect the enrollment data are the quality education offered by UE, very good facilities of UE, good image of UE and affordable tuition fee.

Table 3. Result of Survey – Factors influencing student enrollment in UE

Factors	Year						Mean	Rank
	2010	2011	2012	2013	2014	2015		
A	2719	2959	2161	2259	2422	2539	**2510**	1st
B	1784	1916	1738	1800	1882	1971	**1849**	2nd
C	1254	1482	1529	1626	1509	1274	**1446**	4th
D	758	746	283	493	285	736	**550**	7th
E	1076	1248	615	627	673	1179	**903**	5th
F	968	1132	404	479	729	945	**776**	6th
G	1699	1813	1668	1787	1746	1669	**1730**	3rd
H	131	213	48	109	41	81	**104**	11th
I	59	102	31	61	27	73	**59**	12th
J	396	602	233	211	157	696	**383**	9th
K	580	627	262	256	357	884	**494**	8th
L	135	336	174	139	57	139	**163**	10th

Legend:

A – Quality Education offered by UE
B – Very good facilities of UE
C – Affordable tuition fee
D – Proximity to residence
E – Relative/Friend studied at UE
F – Security
G – Good image of UE
H – Late Enrolment
I – Did not qualify for admission in other school
J – School where I came from was visited
K – Good advertisement
L – Personal choice

B. Design of a polynomial curve fitting model based on freshmen enrolment report

The polynomial curve fitting model was designed through the given n points in the xy-plane $(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)$ that represent a collection of data and the user are asked to find a polynomial function of degree $n - 1$, the model equation is

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1} \quad (1)$$

whose graph passes through the specified points. If all x-coordinates of the points are distinct, then there is precisely one polynomial function of degree $n - 1$ (or less) that fits the n points as shown in Figure 2.

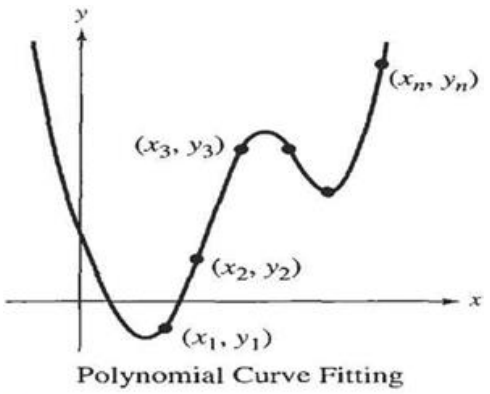

Figure 2. Graph of Polynomial Curve Fitting

The historical data gathered by the researcher was used in the system to predict the enrolment for the coming years. The data was also used to generate the recommender system to be used by the HEI decision maker.

Figure 3 shows the chart for the fetched values for UE Caloocan enrolment data for CFAD from school year 2010 to 2015. Based on the given data, the polynomial function generated using the polynomial curve fitting technique is $p(x) = 152 - 47.93x + 72.42x^2 - 32.17x^3 + 6.08x^4 - 0.4x^5$. Since there are 6 sets of data given, the polynomial function formulated is of 5th degree. It was

P-ISSN 2350-7756 | E-ISSN 2350-8442 | www.apjmr.com
Asia Pacific Journal of Multidisciplinary Research, Vol. 6, No. 4, November 2018 Part II

revealed that the predicted value for the next school year is 297.

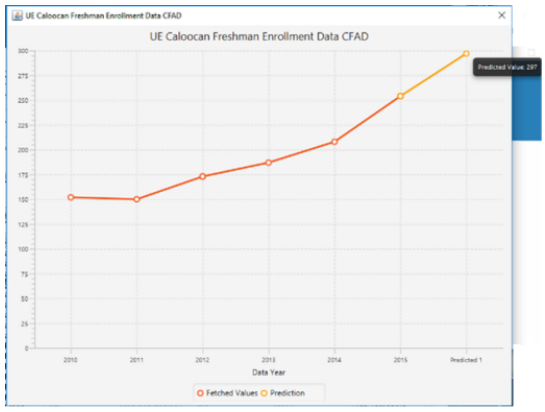

Figure 3. Screenshot of the system showing the chart for UE Caloocan Enrolment Data of the College of Fine Arts, Architecture and Design (CFAD)

C. Efficiency of the forecasting model as basis to develop the recommender system

To effectively evaluate the polynomial curve fitting model, a comparison was done with the other time series forecasting technique like the Simple Moving Average and Exponential Smoothing. Discussed are methods for projecting forthcoming policies.

1. Simple Moving Average

In forecasting the simplest technique to use is the simple moving average. It is done by getting the sum of the number of *n* values and then dividing the sum by total number of *n*. It is the forecast for next period. Its equation is shown in Equation 2 below

$$F_t = \frac{sum\ of\ last\ n\ values}{n} \qquad (2)$$

2. Exponential Smoothing

Exponential smoothing is a renowned technique to produce a smooth out time series. It provides diminishing weighs to observation as it gets older. Equation 3 shows the formula to compute for the next forecasted value. It involves an alpha value (α) that is considered as the smoothing constant. The greater the value of alpha, the nearer to the actual data points and vice versa. This technique is appropriate for projecting figures with no trend or cyclical pattern. To compute for the next predicted value:

$$F_t = F_{(t-1)} + \alpha\,(A_{(t-1)} - F_{(t-1)}) \qquad (3)$$

Table 4 shows the comparison of three forecasting techniques namely the Simple Moving Average, Exponential Smoothing and the Polynomial Curve Fitting Technique for UE Caloocan and UE Manila enrollment data with respect to its Mean Absolute Deviation (MAD), Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE).

MAPE was computed using the formula:

$$|\%Error| = \frac{|Error|}{Actual}\ x\ 100\% \qquad (4)$$

Table 4. Table of comparison of 3 Forecasting Techniques for UE Caloocan and UE Manila Enrolment Data

UE CALOOCAN ENROLLMENT DATA									
College	SMA			EXPONENTIAL SMOOTHING			PCF		
	MAD	MSE	MAPE	MAD	MSE	MAPE	MAD	MSE	MAPE
CAS	27.33	1011.78	8.37%	54.61	4093.30	14.56%	99.20	14945.20	27.70%
CBA	36.11	1605.96	5.91%	34.70	1774.25	6.01%	101.40	14521.40	16.88%
COE	81.78	7432.52	8.48%	51.04	3642.82	5.46%	57.40	4309.40	6.22%
CFAD	43.78	2149.19	19.69%	29.05	1234.56	13.65%	31.80	1317.80	15.38%
Mean (MAPE)			**10.61%**			**9.92%**			**16.55%**
UE MANILA ENROLLMENT DATA									
College	SMA			EXPONENTIAL SMOOTHING			PCF		
	MAD	MSE	MAPE	MAD	MSE	MAPE	MAD	MSE	MAPE
CAS	30.33	1333.07	3.41%	48.69	3802.10	5.80%	76.00	7574.00	9.00%
CBA	53.22	4276.33	5.13%	70.26	5737.53	7.20%	77.20	12636.80	7.77%
CCSS	82.44	6938.00	14.56%	92.48	11020.64	17.52%	116.80	24715.60	20.90%
DENT	27.89	897.96	3.35%	33.14	1278.35	3.96%	96.00	15858.00	11.49%
EDUC	7.11	73.70	7.78%	10.53	155.34	12.03%	18.80	601.20	20.90%
GS	18.56	344.93	17.53%	17.71	389.69	15.71%	29.80	1073.40	27.38%
LAW	28.78	1301.52	26.87%	22.14	700.81	26.71%	9.20	110.80	14.53%
COE	19.22	418.41	2.60%	26.46	1071.02	3.54%	117.40	20635.80	15.66%
Mean (MAPE)			**13.70%**			**14.50%**			**19.62%**

The table above shows the average of MAPE for both UE Caloocan and Manila enrollment data using the three forecasting techniques. It was revealed that polynomial curve fitting technique has a high error as compared to other time series forecasting technique. It only means that the reliability of predicting the enrollment data using polynomial curve fitting is less than the simple moving average and exponential smoothing.

Figure 4. Screenshot of the system to View the Major Factor of UE Caloocan Freshmen Enrolment Data

Figure 4 shows the module of the system to view the major factor of UE Caloocan freshmen enrolment data. It shows that the highest factor or highest predicted value in terms of enrolment is the College of Engineering (COE) and the lowest factor or the lowest predicted value is the College of Fine Arts Architecture and Design (CFAD). It only implies that COE has the highest predicted number of freshmen enrolment and CFAD is the lowest in terms of enrolment result for the coming school year.

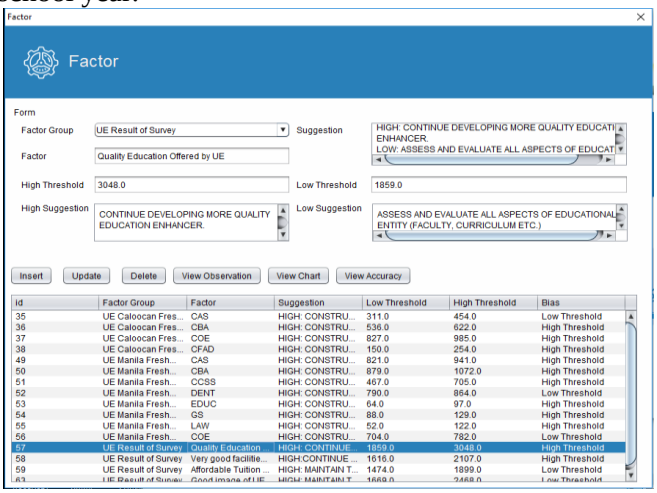

Figure 5. Screenshot of the Recommender system

Figure 5 shows the module for recommender system. It has three suggestions namely the high, low and the general suggestions. The high suggestions may take into considerations as the possible recommendations whenever the predicted value is approaching the high threshold. On the other hand, the low suggestions may take into considerations as the possible recommendations whenever the predicted value is approaching the low threshold. The general suggestion can be a buffer item whenever the user needs it. The last column in the module, labelled 'Bias' will determine the decision to be made by the user.

CONCLUSION AND RECOMMENDATION

The design of a Polynomial Curve Fitting Model has proved different suppositions and possibilities as enumerated:

1. Intensive literature review as well as fact finding technique such as interview and consultation helped the research to identify the predictors significant to the prediction of freshmen enrolment data. Using data mining, it can test and validate the significance of each attributes. On this type of datasets Information Gain Attribute Evaluation using Ranking Filter, logistic regression with conditional forward, and mean was used which produced significant results.
2. The usage of Polynomial Curve Fitting model paved way to an interactive model of information system with much focus on undergraduate freshmen enrolment forecast. The model generated can be translated into computer programs to create a custom prediction model. This research utilized the Polynomial Curve Fitting Technique to forecast the freshmen enrolment data.
3. The different forecasting techniques are essential factors for examining the consequences and vital to be used in future studies. Therefore, the ability to comprehend forecasting methods is significant in determining the best and appropriate tactic that can be a factor of well-informed evidence especially in HEI. The Polynomial Curve Fitting Model can also be used to forecast enrolment data as well as the factors that affect student enrolment but comparing it with other time series forecasting techniques, the Mean Absolute Percentage Error (MAPE) is high.

The design of a Polynomial Curve Fitting Model is open for further improvements; its flexibility in content management from target users per designed by the researcher. Also, it is recommended to:

1. The system could be modified to allow various major or programs and it's the prediction facility though hard coded but displays flexibility based on the capture data model.
2. Formulate a model capturing data from the actual transactions from this system to generate another model and use different algorithms for testing.
3. Additional statistical methods may be combined to the current model to improve the precision of the model. Furthermore, a mixture of the various statistical models can be considered. One form of mixture is implementing cohort survival models of each component of the system and utilized it as an independent variable in projecting the number of students to enrol. Quantitative models can also be delved into by utilizing the factors recognized in this study.

REFERENCES

[1] Ward, J. (2007). Forecasting Enrolment to Achieve Institutional Goals. *College and University Journal*, 82 (3), pp. 41-46.

[2] Mahsood S., & Chenicheri, N. (2010). Enrolling in Higher Education: The Perception of Stakeholders. *Journal of Institutional Research*, pp. 9-15

[3] Bringula, R. & Basa, R. (2010). Institutional image indicators of three Universities: basis for attracting prospective entrants.

[4] Roper, A. T., Cunningham, S., Porter, A., Mason, T., Rossini, F., Backs, J. (2011). Forecasting and Management of Technology. New Jersey: John Wiley & Sons.

[5] Sweeney, S. H., & Middleton, E. J. (2005). Multiregional Cohort Enrolment Projections: Matching Methods to Enrolment Policies. *Population, Space and Place*, 11(5), pp. 361-380.

[6] Nurlida, I., Faridah, H., & Nooraimi, M. (2013). International Students Satisfaction Formation - Linkage between Information Satisfaction and Choice Satisfaction: Study on Private Higher Education Institutions in Malaysia. *AFBE Journal*, pp. 88-108.