



Forecasting the Sum of New College Students with Linear Regression Approach

Yulia Utami^{1,*}, Desi Vinsensia², Aura Nissa³, Sulastri⁴

^{1,3,4} Teknik Informatika, STMIK Pelita Nusantara, Jalan Iskandar Muda No 1 Medan, 20154, Indonesia

² Manajemen Informatika, STMIK Pelita Nusantara, Jalan Iskandar Muda No 1 Medan, 20154, Indonesia

Email: yuliautami14071990@gmail.com¹, desivinsensia87@gmail.com², auranissag@gmail.com³, sulastri130902@gmail.com⁴

* corresponding author

ARTICLE INFO

Article history:

Received: Feb 12, 2022

Revised: Feb 26, 2022

Accepted: Mar 12, 2022

Available online Mar 30, 2022

Keywords:

Forecasting;

Linear regression model;

Mean absolute percentage Error (MAPE).

ABSTRACT

Forecasting models are the result of developments in the field of science and technology that provide convenience in predicting future events. This paper aims to develop a linear regression model to predict the number of new students in the next year. The data to be used in this study is the total of students majoring in informatics engineering and information management during the last 5 years. Based on result obtained the number of student for department of Informatics Engineering is 198 people with a MAPE (Mean Absolute Percentage Error) score of 16.5%, and for the new students department of Informatic Management is 8 people with a MAPE score of 16.1%.

© 2022 JTI C.I.T. All rights reserved.

1. Introduction

A college is a unit of advanced institutions of secondary education aimed at college student in order to become part of a member of society who has the academic ability to engage in innovation in science [1]. College development is influenced by prospective college students in going through education in college, great interest candidate college students is customized by human resources, facilities and infrastructure [2]. In the learning process activities are influenced by the sum of college students admitted because one of the things used in planning during the course is college students. Sum of college students admitted is related to the sum of lecturers, as well as the facilities a college [3]. Therefore, a method for forecasting required in predicting the sum of college students who will register in a college. Forecasting is part of an effective and efficient planning process for private colleges in maximizing human resources and increase quality of service [4]. Income a private college relies on by college student payments, so forecasting of the sum of college students is very important for private colleges [5]. Forecasting is said to be a calculation for determining an outcome or an impending event with a mathematical model, and in its calculation it can be done either with quantitative or qualitative approaches [6]. Qualitative prediction are difficult to do to get good result because of the variable is very relative [7]. Quantitative prediction result depend the method used [8]. According Ginting, forecasting is a demand in the future, which is basically an estimate (guess), but with certain techniques forecasting can be better than a regular estimate so that forecasting can be declared a scientific estimate [9]. In this research the forecasting method used is a simple linier regression method. Simple linier regression method is a statistical method that serves to test the extent of the causal relationship between the predictor (independent) variable (x) and corrector variable (y) [10]. Linear regression method has been widely used in various fields as one of simple predicting technique such as: education to predict number of final grade students [11], forecasting academic performance in military environment [12]. Forecasting future economic national model [13], [14], [15], prediction result of business [16], [17], prediction number of visiting patient [18],



[19], etc. The purpose of this study is to predict the number of new student admissions to college in the next year to maximize profits by using a simple linear regression model approach. In this study consider the last 5 years academic period to predict the number of student in the next year for two major in college and using mean absolute percentage error (MAPE) to determine the level of accuracy the predicted value. This research has 4 parts, among others: part 2 describes the research method and present with a fishbone diagram. Section 3 provides the results and discussion. Section 4 summarizes the conclusions from the discussion carried out in the previous section.

2. Research Method

The purpose of this research is to predict the number of students in the departments of informatics engineering and informatics management who will register in the STMIK Pelita Nusantara campus in the following year. After learning the results of the forecasting, then the college can prepare some planning and strategies in acquiring prospective new students. This study's primary data taken as independent variables are from the previous academic year during 5 years from 2017 until 2022.

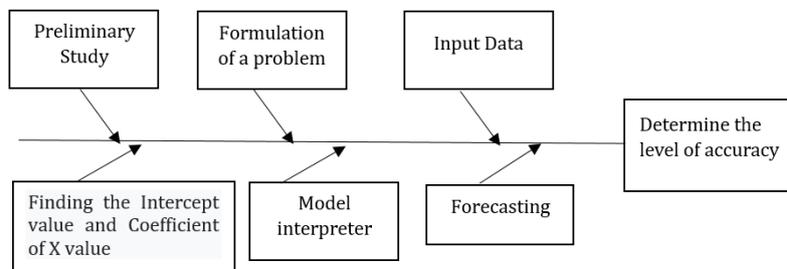


Fig 1. Fishbone Diagram

The diagram at figure 1 explain the steps of flows this research, as follows:

- The initial stage in this research is to identify problems related to linear regression and forecasting and collect data.
- The next step is to determine the problem formulation, where the problem formulation in this study is to determine a simple linear regression model based on the data for the last 5 years and to prediction the number of new students in the next year.
- Input data. In this step, using as input data total number of student for last 5 years.
- Find the intercept value (a) and coefficient of x value (b), the following steps is count X^2 , Y^2 , XY and sum of each [18], than enter to formula

$$a = \frac{(\sum X^2)(\sum Y) - (\sum XY)(\sum X)}{n(\sum X^2) - (\sum X)^2} \quad (1)$$

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2} \quad (2)$$

In this steps, using SPSS application to determine value (a) and (b).

- Substituted equations 1 and 2 to the formula linear regression
- $$Y = a + bX \quad (3)$$

Such as

Y = dependent variable

a = intercept

b = coefficient of X

X = independent variable

- The next steps is forecasting. Forecasting is prediction by used certain technique. The following to find forecasting is:
 - Determine the problems to be analyzed and collect the required data
 - Prepared data
 - Establish and apply methods

4) Make predictions on data for the future

5) Evaluate forecasting results

This researc^h, forecasting sum new students who will register the new academic years can be found by replacing the variable $x = 6$

- g. MAPE testing in linear regression is used to determine the level of accuracy in the calculation. To determine the degree of accuracy, can use MAPE method by calculating the difference from actual data and prediction data [19]. The smaller the MAPE value then the higher the degree of accuracy, than other the better the method used in forecasting [20]. In determining the MAPE value, equation

$$MAPE = \frac{\sum \frac{|Y-Y'|}{Y}}{n} \times 100\% \tag{4}$$

Such as:

Y = actual data

Y' = prediction data

n = sum of data

3. Results And Discussion

The important thing to consider regression analysis is that underlying relationships among the variables cannot be determined. While the terminology is such that we say that X "predicts" Y, we cannot say that X "causes" Y [14]. Dependent variable is effect variable or affected variable, where as independent variable is cause variable or variables affect [15]. There are two variables in this research, namely the independent variable symbolized by x represent as the academic years, and the dependent variable symbolized by Y represent as the number of new students. To find the value of the intercept and coefficient X, using the SPSS software. The regression equation can be arranged with the Intercept value and the coefficient of the X variable in column B and the sig. of the X variable value is analyzed.

Table 1.
Intercept Value And Coefficient X Value

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	T	Sig.
1	(Constant)	307.600	68.295		4.504	.020
	tahun ajaran	-18.800	20.592	-.466	-.913	.429

a. Dependent Variable: Number of IT Students

The table 1 shows that the value of regression equation $Y = 307.6 - 18.8 X$, with constant value is 307.6. That results means, based on the calculation statistics without X, value of Y is 307.6 or 308 students. Value Y represented that the number of students for department of informatics engineering (IT) without X value is 308 students.

Table 2.
R square value

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.466a	.217	-.043	65.11631

a. Predictors: (Constant), academic year

R Square is a value that shows the percentage of the contribution of the independent variable in influencing the dependent variable while the rest is influenced other variables or epsilon which is symbolized ϵ_i . The table 2 shown value R Square is 0.217, which means the contribution of the independent variable in influence the dependent variable is 21.7% while the remaining 78.3% is influenced the other variables.

Table 3.
Intercept value and coefficient x value

Model		Unstandardized Coefficients		Standardized coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	32.700	4.810		6.798	.007
	tahun ajaran	-4.100	1.450	-.853	-2.827	.066

a. Dependent Variable: Number of informatics management Students

Table 3 shown the regression equation for departments of informatics management, where the regression linear equation with $Y = 32.7 - 4.1 X$. That's equation arranged with the Intercept value and the coefficient of the X variable in column B and analyzed the sig the X variable. The intercept value is 32.7, it means that statistics without X so Y is 32.7 or 33 students.

Table 4
R square value

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.853a	.727	.636	4.58621

a. Predictors: (Constant), academic year

R Square is a value that shows the percentage of the contribution of the independent variable in influencing the dependent variable while the rest is influenced by other variables or epsilon which is symbolized ϵ . Table 4 shown that value of R Square for department of informatics management is 0.727. This means that the contribution of the independent variable in influencing the dependent variable is 72.7% while the remaining 29.3% is influenced by other variables.

Table 5.
Linier regression equation and forecasting

Department	Linier Regression Equation	Forecasting
Department of Informatics Engineering	$Y = 307,6 - 18,8 X$	$Y = 307,6 - 18,8(6) = 194,8 = 195$
Department of Informatics Management	$Y = 32,7 - 4,1 X$	$Y = 32,7 - 4(6) = 8,1 = 8$

Using data the number of students from previous 5 years academic yaeer with linear regression methods obtained for model linear regression of department of informatics engineering and department of informatics managemnt as shown on the table 5. In addition on the table 5, also shows the results of prediction the number of students for each major. Table 6 presented data the number of students from 2017 until 2021 and also shows the result of prediction number of students in academic year 2021 until 2022. Table 7 present the result of different prediction the number of students from 2017 until 2022. After determined the linear regression model, then calculate value of mean average percentage error (MAPE) or called MAPE test for each linear regression model of major.

Table 6
Calculation of predictive data (y')

Academic Years	Department of Informatics Engineering	Department of Informatics Management
2017 - 2018	289	29
2018 - 2019	270	25
2019 - 2020	251	20
2020 - 2021	232	16
2021 - 2022	214	12

Table 7.
Prediction difference calculation ($y - y'$)

Academic Years	Department of Informatics Engineering	Department of Informatics Management
2017 – 2018	-54	4
2018 – 2019	14	-6
2019 – 2020	86	1
2020 – 2021	2	-2
2021 – 2022	-48	3

Table 8
Mape error rate calculation

Academic Years	Department of Informatics Engineering	Department of Informatics Management
2017 – 2018	0,229	0,133
2018 – 2019	0,049	0,289
2019 – 2020	0,255	0,029
2020 – 2021	0,007	0,164
2021 – 2022	0,287	0,187
total	0,826	0,802
MAPE	16,5	16,1

Table 8 shows that by using the MAPE formula for Department of Informatics Engineering of 16.5% and MAPE values for Department of Informatics Management of 16.1%. Therefore it means, the MAPE values of each major can be categorized as good.

4. Conclusion

Based on the results of the discussion, this study can be used as prediction the new student admissions in the coming year based on data from the previous period. This study using the data previous of 5 years the results obtained the equation $Y = 307.6 - 18.8 X$ for department of infomatics engineering and the equation $Y = 32.7 - 4.1 X$ for department of informatics management. In addition, the accuracy of the calculations carried out in forecasting the number of new students has resulted in an error value of 16.5% for department of informatics engineering and 16.1% for department of informatics management using the MAPE model. For future study, will be consider to adding independent variables to the forecasting criteria or combining them with other methods to get closer to the real result.

References

- [1] Y. S. Bhakti, A. B. Kusdinar, dan A. Sunarto, "Model Peramalan Penerimaan Calon Mahasiswa Menggunakan Metode Regresi," hal. 113–120.
- [2] S. A. Alven Safik Ritonga, "Pengembangan Model Jaringan Syaraf Tiruan untuk Memprediksi Jumlah Mahasiswa Baru di PTS Surabaya (Studi Kasus Universitas Wijaya Putra)," vol. 12, no. 1, hal. 15–24, 2018.
- [3] N. Almutzah, N. Azizah, Y. L. Putri, I. Negeri, dan S. Ampel, "Prediksi jumlah mahasiswa baru menggunakan metode regresi linier sederhana," vol. 18, hal. 31–40, 2021.
- [4] A. Bengnga dan R. Ishak, "Prediksi Jumlah Mahasiswa Registrasi Per Semester Menggunakan Linier Regresi Pada Universitas Ichsan Gorontalo," *Ilk. J. Ilm.*, vol. 10, no. 2, hal. 136–143, 2018, doi: 10.33096/ilkom.v10i2.274.136-143.
- [5] S. A. Yasinta dan Y. Yudihartanti, "Prediksi Daya Tersambung Dengan Metode Double Exponential Smoothing," *Jutisi*, vol. 5, no. 2, hal. 1151–1162, 2016.
- [6] T. N. Putri, A. Jordan, dan D. H. Lamkaruna, "Peramalan Penerimaan Mahasiswa Baru Universitas Samudra Menggunakan Metode Regresi Linear Sederhana," *J-TIFA (Jurnal Teknol. Inform.*, vol. 2, no. 1, hal. 21–27, 2019.
- [7] M. Syafruddin, L. Hakim, dan D. Despa, "Metode Regresi Linier Untuk Prediksi Kebutuhan Energi Listrik Jangka Panjang (Studi Kasus Provinsi Lampung)," *J. Inform. dan Tek. Elektro Terap.*, vol. 2, no. 2, 2014, doi: 10.23960/jitet.v2i2.237.

- [8] M. Hakimah, R. R. Muhima, dan A. Yustina, "Persediaan Barang Dengan Metode Trend Projection," *SimanteC*, vol. 5, no. 1, hal. 37–48, 2015.
- [9] S. R. Febri Andrian, Shantika Martha, "Sistem peramalan jumlah mahasiswa baru menggunakan metode," vol. 08, no. 01, hal. 112–121, 2020.
- [10] D. A. Trianggana, "Peramalan Jumlah Siswa-Siswi Melalui Pendekatan Metode Regresi Linear," *J. Media Infotama*, vol. 16, no. 2, hal. 115–120, 2020, doi: 10.37676/jmi.v16i2.1149.
- [11] H. Darman, S. Musa, R. Ramasamy, dan R. Rajeswari, "Predicting students' final grade in mathematics module using multiple linear regression," *Int. J. Recent Technol. Eng.*, vol. 7, no. 5, hal. 331–335, 2019.
- [12] W. Z. Khan dan S. Al Zubaidy, "Prediction of Student Performance in Academic and Military Learning Environment: Use of Multiple Linear Regression Predictive Model and Hypothesis Testing," *Int. J. High. Educ.*, vol. 6, no. 4, hal. 152, 2017, doi: 10.5430/ijhe.v6n4p152.
- [13] Y. Xiao dan Z. Jin, "The Forecast Research of Linear Regression Forecast Model in National Economy," *OALib*, vol. 08, no. 08, hal. 1–17, 2021, doi: 10.4236/oalib.1107797.
- [14] S. F. Shazmeen, M. M. A. Baig, dan M. R. Pawar, "Regression Analysis and Statistical Approach on Socio-Economic Data," *Int. J. Adv. Comput. Res.*, vol. 3, no. 3, hal. 347–351, 2013.
- [15] G. N. Ayuni dan D. Fitriana, "Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ," *J. Telemat.*, vol. 14, no. 2, hal. 79–86, 2019.
- [16] J. Rusov, M. Misita, D. D. Milanovic, dan D. L. Milanovic, "Applying regression models to predict business results," *FME Trans.*, vol. 45, no. 1, hal. 198–202, 2017, doi: 10.5937/fmet1701198R.
- [17] M. A. Maricar et al., "Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ," *J. Telemat.*, vol. 10, no. 2, hal. 79–86, 2019, [Daring]. Tersedia pada: <https://www.jurnal.umk.ac.id/index.php/simet/article/view/3484>.
- [18] W. M. Baihaqi, M. Dianingrum, dan K. A. N. Ramadhan, "Regresi Linier Sederhana Untuk Memprediksi Kunjungan Pasien Di Rumah Sakit Berdasarkan Jenis Layanan Dan Umur Pasien," *J. Simetris*, vol. 10, no. 2, hal. 671–680, 2019.
- [19] C. V. Hudyanti, F. A. Bachtiar, dan B. D. Setiawan, "Perbandingan Double Moving Average dan Double Exponential Smoothing untuk Peramalan Jumlah Kedatangan Wisatawan Mancanegara di Bandara Ngurah Rai," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, hal. 2667–2672, 2019.
- [20] M. A. Maricar, "Analisa Perbandingan Nilai Akurasi Moving Average Dan Exponential Smoothing Untuk Sistem Peramalan Pendapatan Pada Perusahaan XYZ," *J. Sist. dan Inform.*, vol. 13, no. 2, hal. 36–45, 2019.