Analysis of the Influence and Prediction of the Number of Students on PNBP using Multiple Regression

Alfrianus Papuas
Department of Information System
North Nusa State Polytechnic

Ella H. Israel
Department of Information System
North Nusa State Polytechnic

Noldy Sinsu
Department of Information System
North Nusa State Polytechnic

ABSTRACT
Conduct forecasting analysis with good accuracy for other PNBP components which can ultimately be the basis of PNBP receipt projection. Multiple Linear Regression is used to predict how the state (ups and downs) of independent variables, when two independent variables as predictor factors are manipulated (the ups and downs of values). By analyzing the relationship between several variables to measure the degree of relationship and the direction of the relationship between independent variables and dependent variables then predict a variable to forecast in the future. In this study simultaneously independent variables namely The Number of Student and Facilities influenced dependent variables namely PNBP and through multiple linear regression equations produced by future forecasting with the results of measurement of forecasting errors are very small.

General Terms
Multiple Linier Regression

Keywords
Multiple Linier Regression, Prediksi, PNBP

1. INTRODUCTION
Non-Tax State Revenue Management as known as Penerimaan Negara Bukan Pajak (PNBP) is the utilization of resources in the framework of governance which includes planning, implementation, accountability, and supervision activities to improve services, accountability, and optimization of state revenues derived from Non-Tax PNBP [1]. The PNBP type of the Ministry of National Education is regulated in Government Regulation No. 22 of 1997 concerning Types and Deposits of Non-Tax State Revenues, one of which is The Receipt of education implementation [2]. The importance of the effectiveness of PNBP management is related to planning, the achievement of targets, and realization of budgets and quality of services to the community [3]. Conducting projection/forecasting analysis with good accuracy for SDA revenue components, SOE profit, and other PNBP components that can ultimately become the basis of PNBP revenue projection (Permenkeu No. 100/2018 Pasal 1825).

Politeknik Negeri Nusa Utara (Polnustar) is one of the educational institutions in the archipelago whose financing is sourced from the State Budget (APBN) and Non-tax State Revenue (PNBP). Good management of PNBP continues to be carried out to achieve the target in the coming year, which is a component of PNBP acceptance in Polnustar sourced from education cost revenue, other education income, and rental of existing facilities. For the source of income through education costs obtained from a single tuition fee whose amount of admission depends on the number of students while PNBP sourced from the rental of facilities, the management is based on the type and rates.

In this study, we want to analyze which free variables have a dominant influence on PNBP and how much influence those variables on PNBP and forecast for future PNBP acceptance. By using multiple regression analysis to examine the relationship of multiple variables and predict a variable and perform prediction error rate calculations using Mean Absolute Deviation (MAD), Mean Squared Error (MSE), and Mean Absolute Percent Error (MAPE) where the accuracy of forecasting will be higher when the values of MAD, MSE, and MAPE are smaller [4].

2. METHODOLOGY
2.1. Data Collection
Metode yang digunakan untuk studi ini adalah regresi berganda. Regresi Linear Berganda digunakan untuk meramalkan bagaimana keadaan (naik turunnya) variabel independen, bila dua variabel indevenden sebagai faktor predictor dimanipulasi (naik turunnya value ofi) [5].

The data used is secondary data, namely data on the number of PNBP receipts and facility rental data obtained from Polnustar finance and student count data obtained from Polnustar academic section for 7 years.

Tabel 1. PNBP Raw Dataset 2014-2020

Tahun	PNBP	MHS	FLS
2014	4.193.732.283	854	117.207.700
2015	3.461.985.441	956	287
2016	4.707.096.033	747	9.084.242
2017	3.590.216.453	860	213.157.002
2018	3.237.565.566	893	128.004.340
2019	2.942.123.734	616	168.686.340
2020	2.498.220.840	846	78.162.840

From the existing data, it can be seen that the data is still not able to be done data processing because there is still a very long range and can cause overfitting. Therefore, raw data processing is required so that it can be used at the prediction stage. In this research, we use the log of each variable, the result can be seen in Table 2.

Tabel 2. PNBP Log Dataset 2014-2020

Tahun	PNBP	MHS	FLS
2014	9,602	2,931	8,069
2015	9,539	2,980	2,458
2016	9,673	2,942	6,958
2017	9,555	2,876	8,329
2018	9,510	2,756	8,107
2019	9,469	2,712	8,227
2020	9,398	2,695	7,893

2.2. Research Step

The stages that will be carried out are problem identification, data collection, parameter formation, parameter data processing used in the calculation process using Multiple Linear Regression.

2.2.1 Multiple Linear Regression

Regression analysis is one of the data analysis techniques in statistics that are often used to examine the relationship between multiple variables and predict a variable [6]. If you want to review the relationship or influence of one free variable on a non-free variable, then the regression model used is a simple linear regression model. Then If you want to review the relationship or influence of two or more free variables on non-free variables, the regression model used is a multiple linear regression model. The general form of multiple linear regression models with free variables is as in the following equation (2.1) [6].

$$Y_i = \beta_0 + \beta_1 X_{i\,1} + \beta_2 X_{i\,2} + \cdots + \beta_{p-} X_{i\ ,p\,-1} + \varepsilon_i$$

With :

Y_i is a non-free variable for i-observation, for i = 1,2,…,n.

$\beta_0, \beta_1, \beta_2, \ldots, \beta_p - 1$ is a parameter.

$X_{i\,1}, X_{i\,2}, \ldots, X_{i\ ,p\,-}\,1$ is a free variable

Ɛ is the residual (error) for observations that are assumed to be normal distributions that are mutually free and identical to the average of 0 (zero) and variance. The partial parameter testing procedure is as follows:

1. Hypothesize
   $H_0 = \beta_\kappa = 0$
   $H_1 = \beta_\kappa \neq 0$, for k = 1, 2, …., p-1.
   (Kutner, et.al., 2004)
   or :
   H_0 : The k-free variable does not affect the free variable
   H_1 : K-free variable affects non-free variable for k = 1, 2,…, p-1
2. Determine a significant level (a)
   The significant (α) rate often used in research is 5%
3. Determine test statistics
   The test statistics used are:
   $$t = \frac{b_k}{s\,(b_k)}$$
   With :
   b_k is the estimated value of the β_κ (obtained from the OLS method).
   $s\,(b_k)$ is the standard deviation of the estimated parameter value β_κ
4. Determining the criticism used
5. Draw conclusions

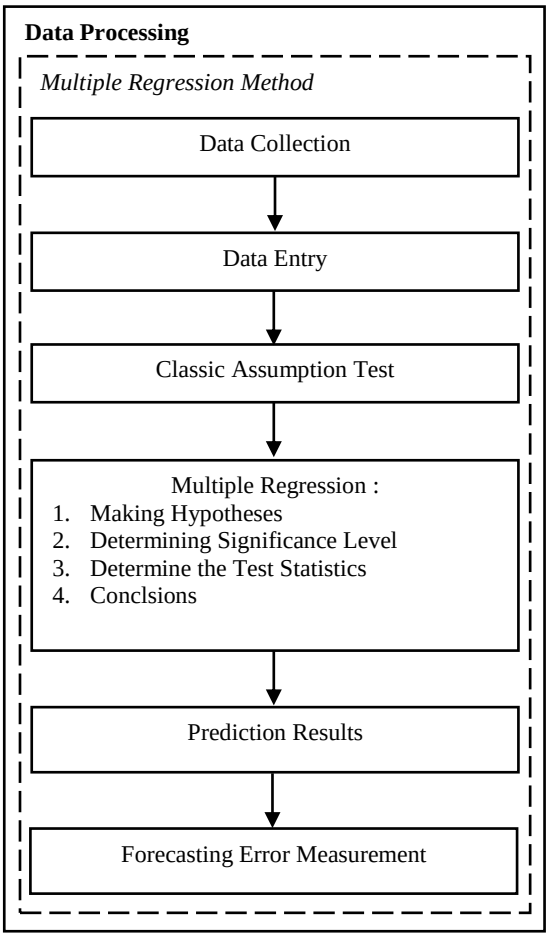

Fig 1. Multiple Regression Method

2.2.2 Forecasting Error Measurement

There are three calculation methods commonly used to calculate forecasting errors among others, as follows [7] :

1. Mean Absolute Deviation (MAD)
   Mean absolute deviation is the first measure of an overall forecasting error for a model. This value is calculated by taking the sum of the absolute values of each forecasting error divided by the number of data periods (n). The MAD value is calculated by the following formula:
   $$\text{MAD} = \frac{\Sigma\,(At - Ft)|}{n}$$
2. Mean Squared Error (MSE)
   Mean Squared Error is the second way to measure overall forecasting errors. MSE is the average difference of squares between predicted and observed values. Generally the smaller the MSE value, the more accurate the forecast. MSE values are calculated by the following formula::
   $$\text{MSE} = \frac{\Sigma(At - Ft)^2}{n}$$
3. Mean Absolute Percent Error (MAPE)

Mean Absolute Percent Error is calculated as the average absolute difference between predicted and actual values, expressed as a percentage of the actual value. Mape values are calculated by the following formula:

$$\text{MAPE} = \frac{\sum (At - Ft) / At}{n} \text{x}100\%$$

3. RESULT AND ANALYSIS

To achieve the research objectives discussed earlier, several stages must be done. The stages are divided into several processes including, data collection, before entering into the process of measuring forecasting errors that are part of data modeling to determine the multiple regression equations that will be used in the process of prediction results for the application of the Multiple Regression method of making hypotheses, determining significant levels, determining test statistics and drawing conclusions.

3.1. Application of *Multiple Regression*

The data analysis techniques in this study used multiple linear regression analysis assisted by SPSS version 26, to obtain a comprehensive picture of the relationship between variables. Before multiple linear regression analysis, several assumptions must be met, namely Normality Test, Multicolerity Test, Heteroskedastisity Test, and Autocorrelation Test. The results of application in this study using Model (MDL), Inserted Variable (VARS+), Issued Variable (VARS-), and Method used can be seen in Table 3.

Tabel 3. Input/Elimination Variable [a]

MDL	VARS+	VARS-	Method
1	FLS, MHS[b]	-	Enter

The independent variable in this analysis is the variable number of students and facilities, while the dependent variable is Penerimaan Negara Bukan Pajak (PNBP), no variable is discarded so that in the variables Removed column there are no numbers or blanks.

3.1.1 Normality Test

In the normal P-P Plot chart above, residual data dissemination can already be said to simply follow the normal line (straight line). The results can be seen in Figure 2.

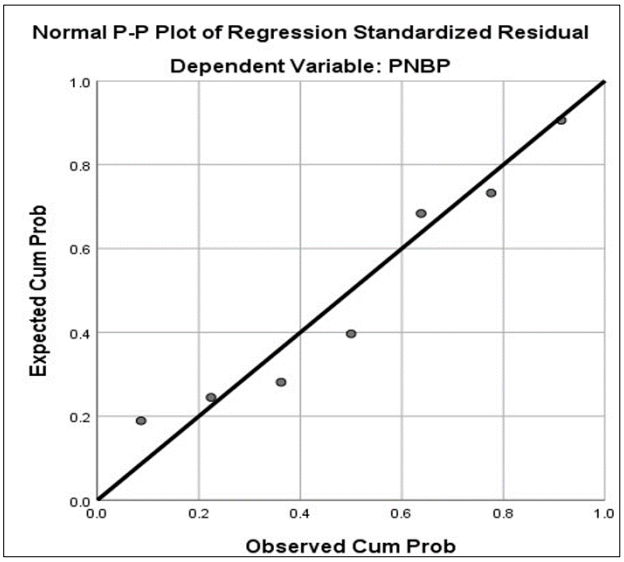

Fig 2. Normal Probability Plot

3.1.2 Multicolerity Test

Multicolerity is a situation where some or all of the free variables are strongly correlated. Thus, the greater the correlation between fellow independent variables, the greater the error rate of the regression coefficient resulting in the higher the default error. The way used to detect the presence or absence of multicoliniearity is to: use Variance Inflation Factors (VIF) or can be seen also from the tolerance value as shown in Table 4.

Tabel 4. Statistics of Kolinierity

Tolerance	VIF
,660	1,514
,660	1,514

it can be concluded that there is no multicolerity because the tolerance value is already greater than 0.1 and the VIF value is less than 10. Thus there is no strong relationship between the Variable Number of Students and Facilities.

3.1.3 Heteroscedastisity Test

Heteroscedastisity is a condition in which in the regression model there is variance inequality from residual one observation to another. If the variance from residual one observation to another remains, then it is called homoscedastisitas and if the variance is different it is called heteroscedstisity. In this study will be used chart method (scatterplot chart) to test heteroskedasticity.

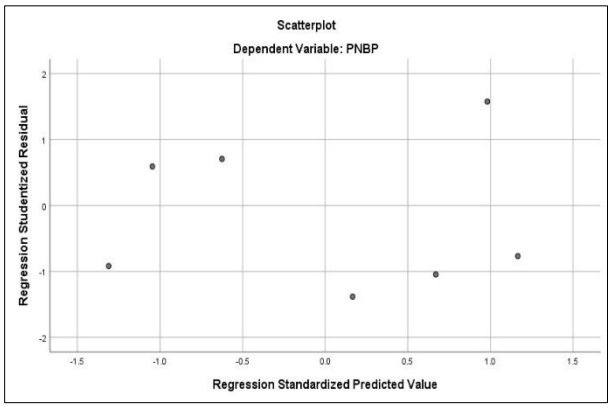

Fig 3. Heteroscedasticity Test Results

Heteroscedasticity is a condition in which in the regression model there is variance inequality from residual one observation to another. If the variance from residual one observation to another remains, then it is called homoscedasticity and if a variance is different it is called heteroscedasticity. In this study will be used chart method (scatterplot chart) to test heteroskedasticity.

3.1.4 Autocorrelation Test

Autocorrelation or self-correlation or serial correlation is a condition in which there is a correlation between the sequential values of the same variable. In this study, an autocorrelation test was conducted with Durbin Watson's statistical test by comparing durbin watson's calculated statistical value with Watson's durbin value in Table 5.

Tabel 5. Autocorrelation Results using Durbin Watson

Change Statistic					
R Sqr Change	**F Change**	**Df1**	**Df2**	**Sig. F. Change**	**Durbin-Watson**
,856	11,865	2	4	,021	2,763

Based on the results of the analysis above it is known that DW = 2763, with Based on weston test decision making that states that if the DW value is between 4-dU (2763) to 4-dL (3,533) then the results can not be concluded, thus it will be tested using run test statistics and the results can be seen in Table 6.

Tabel 6. Autocorrelations Results using Runs Test

Run Test	**Unstandardized Predicted Value**
Test Value[a]	9,54997
Cases < Test Value	3
Cases >= Test Value	4
Total Cases	7
NumberOfRuns	2
Z	-1,637
Asymp. Sig. (2-tailed)	,102

1. If the value is Asymp. Sig. (2-tailed) smaller < than 0.05 then there are symptoms of autocorrelation.
2. If the value is Asymp. Sig. (2-tailed) greater than 0.05 then no autocorrelation symptoms.

Based on the results of the above analysis obtained results that the value of asymp. A sig of 0.102 is greater than 0.05, it can be concluded that there are no autocorrelation symptoms or no strong residual relationship between the models. So the autocorrelation problem that cannot be solved with Durbin Watson can be solved by a test run test.

3.2. Multiple Linear Regression Analysis

By using the help of SPSS program application, obtained the output of multiple linear regression calculation results as follows:

Tabel 7. Multiple Regression Analysis Test Results

Model 1	**Unstandardized Coefficients**		**Standardized Coefficients Beta**	**t**	**Sig.**	**Correlations**			**Colinearity Statistics**	
	B	**Std.Error**				**Zero-Order**	**Partial**	**Part**	**Tolerance**	**VIF**
(Constant)	6,939	,550	-	12,615	,000	-				
MHS	,858	,178	1,127	4,822	,009	,821	,924	,916	,660	1,514
FLS	,022	,010	,525	2,247	,088	-,131	,747	,427	,660	1,514

Based on the above output obtained constant value and coefficient of regression so that can be formed multiple linear regression equations as follows:

Y = 6,939 + 0.858 (Number of Students) + 0.022 (Facilities)

1. The constant value obtained is 6,939, this means that if the Number of Students and Facilities variables are assumed to be equal to zero, then the PNBP value is 6,939.

2. The coefficient value of The Number of Students is 0.858, this means that if the variable Number of Students increases by one point, while the facility variable is considered fixed it will cause an increase in PNBP of 0.858.

3. Facility coefficient value of 0.022, this means that if the Facility variable increases by one point, while the Variable Number of Students is considered fixed it will cause an increase in PNBP by 0.022.

This correlation is used to measure the degree of relationship and the direction of the relationship between independent variables namely Number of Students (X1) and Facilities (X2), with dependent variables of (PNBP). Pendapatan Negara Bukan Pajak (PNBP).

Tabel 8. Correlations

Correlations		**PNBP**	**MHS**	**FLS**
Pearson	PNBP	1,000	,821	-,131
	MHS	,821	1,000	-,583
	FLS	-,131	-,583	1,000
Sig. (1-Tailed)	PNBP	-	,012	,389
	MHS	,012	-	,085
	FLS	,389	,085	-
N	PNBP	7	7	7
	MHS	7	7	7
	FLS	7	7	7

According to Sugiyono (2018)[8] guidelines for providing interpretation of correlation coefficients as follows:
0,00 - 0,199 = very low
0,20 - 0,399 = low
0,40 - 0,599 = medium
0,60 - 0,799 = strong
0,80 - 1,000 = very strong
Based on the table above it is known that PNBP has a significant positive relationship and is categorized very strongly with the Number of Students where the correlation coefficient is 0.821 while PNBP has a negative result and is categorized very low with facilities where the correlation coefficient is -0.131.

To find out the coefficient of determination of SPSS, the output results of the coefficient of determination are as follows :

Tabel 9. Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics				Durbin-Watson
						F Change	Df1	Df2	Sig. F Change	
1	,925[a]	,856	,784	,041897	,856	11,865	2	4	,021	2,763

Based on the above results, it is known that the value of the coefficient of determination or R square is 0.856, this value comes from the squaring of the correlation coefficient value (R) which is 0.925 X 0.925= 0.856. The amount of coefficient of determination is 0.856 or equal to 85.6 %, this figure indicates that the variable Number of Students (X1) and Facility Variable (X2) simultaneously (together) affect the variable PNBP (Y) by 85.6 %, while the remaining 4.4 influenced other variables not observed in this study.

3.2.1 Simultaneous Test (F-Test)

Simultaneous test (F-test) To test the correctness of the hypothesis first used statistical test F, namely to test the meaning of the influence of all free variables (independent) together against bound variables (dependents). Hypothesis testing is formulated as follows:

If the value is Sig. < 0.05, then the hypothesis is accepted, it means the number of students (X1) and facilities (X2) simultaneously affects PNBP (Y)

If the value is Sig. 0.05, then the hypothesis is rejected, meaning the number of students (X1) and facilities (X2) simultaneously does not affect PNBP (Y)

Tabel 10. ANOVA[a]

Model 1	Sum of Squares	df	Mean Square	F	Sig.
Regression	0,42	2	,021	11,865	,021[b]
Residual	,007	4	,002	-	-
Total	,049	6	-	-	-

Based on the results of the above analysis is a known sig value. is 0.021. Because of the sig value. 0.021 < 0.05, then following the basis of decision making in test F can be concluded that the hypothesis can be accepted with this number of students and facilities simultaneously affect PNBP.

If the value of F counts > F table, then the hypothesis is accepted, then it means the number of students (X1) and facilities (X2) simultaneously affects PNBP (Y)
If the value of F count < F table, then the hypothesis is rejected, then it means the number of students (X1) and facilities (X2) simultaneously does not affect PNBP (Y)
Based on the results of the analysis above known value of F count is 11,865. Because the value of F calculates 11,865 > 5.79 which is significant at 5 % or 0.05, then following the basis of decision making in test F can be concluded that the hypothesis can be accepted with this number of students and facilities simultaneously affect PNBP.

The value F table = 5.79, where F table = K;n-k(k is the number of variables and n is the amount of data) (2 ; 7-2) = 2: 5 in the table distibusi value F.

3.2.2 Partial Test (T Test)

1. H1 : there is an influence on the number of students on PNBP
2. H2 : there is an influence of facilities on PNBP

Based on significant value

1. If the value is significant sig. < probability is 0.05 then there is an influence of free variable (X) on the bound variable (Y) or accepted hypothesis.
2. If the value is significant sig. > probability 0.05 then there is no effect of the free variable (X) on the bound variable (Y) or hypothesis rejected.

Based on the comparison of calculated t and t table values

1. If the value of t count > t table then there is an influence of free variable (X) on the bound variable (Y) or hypothesis accepted.
2. If the value of t count < t table then there is no influence of free variable (X) on the bound variable (Y) or hypothesis rejected.

Tabel 11. Coeeficients[a]

Model 1	Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.	Correlations			Colinearity Statistics	
	B	Std.Error				Zero-Order	Partial	Part	Tolerance	VIF
(Constant)	6,939	,550	-	12,615	,000	-				
MHS	,858	,178	1,127	4,822	,009	,821	,924	,916	,660	1,514
FLS	,022	,010	,525	2,247	,088	-,131	,747	,427	,660	1,514

1) Partial t Test = Number of Students
Berdasarkan hasil analisis diatas diketahui value of sig. variable Jumlah Mahasiswa adalah sebesar 0,009. Karena value ofi sig. 0,009 < probabilita 0.05, maka dapat diambil kesimpulan bahwa H1 atau hipotesis pertama dapat diterima dengan ini jumlah mahasiswa berpengaruh secara parsial terhadap PNBP.

Based on the results of the analysis above known value of t count variable Number of Students is 4,822. Because the calculated t value of 4,822 > t table 2,776 can be concluded that H1 or the first hypothesis can be accepted by this number of students partially affect PNBP.
Value t Table = (a/2 ; n-k-1) = (0.05/2 ; 7-2-1) = (0.025 ; 4) distribution value t table = 2.776.

2) Partial t Test = Facility
Based on the results of the above analysis is known value of sig. facility variable is 0.088. Because of the sig value. 0.088 > probability 0.05, it can be concluded that H2 or the Second hypothesis is rejected hereby the facility has no partial effect on PNBP.

Based on the results of the analysis above known value of t calculate variable Facility is 2,247. Because the calculated t value of 2.247 < t table 2.776 can be concluded that H2 or the second hypothesis is rejected hereby the facility has no partial effect on PNBP.
Value t babel = (a/2 ; n-k-1) = (0.05/2 ; 7-2-1) = (0.025 ; 4) distribution value t table = 2.776.

Once the linear regression equation is obtained, then for the prediction of Non-Tax State Revenue can be done. To calculate the predicted value of PNBP in 2021, by entering the values X1 and X2 in the last year period (year 2020), where in 2020 the value of X1 = 2,927 and the value X2 = 7,893, then calculate the prediction of PNBP using the double linear regression equation above, so: Y = a + b1. XI + b2. X2

= 6.939 + 0.858 X1 + 0.022 X2

= 6.939 + 0.858 (2.927) + 0.022 (7.893)

= 6.939 + 2.51 + 0.17

= 9,619

Thus the result of the predicted value of Non-Tax State Revenue in the period 2021 is 9,619 or Rp 4,159,106,105. There was an increase in revenue from 9,398 in 2020 (Rp. 2,498,220,840) to 9,619 (Rp 4,159,106,105) in 2021.

3.3. Measurement of Prediction Errors

The result of calculation of prediction errors by calculating the values Mean Absolute Deviation (MAD), Mean Square Error (MSE), Mean Absolute Percentage Error (MAPE), where the value of MAD = 0.040, MSE = 0.002 and MAPE = 0.419% hereby the resulting value is relatively small, so the model used in this forecasting is good.

4. CONCLUSION

1. In this study obtained the regression equation: Y = 6,939 + 0.858 (Number of Students) + 0.022 (Facility), where the constant value is obtained by 6,939, this means that if the variable Number of Students and Facilities is assumed to be equal to zero, then the PNBP value is 6,939. with a correlation coefficient value shows that PNBP has a significant positive relationship and is categorized very strongly with the Number of Students where the correlation coefficient is 0.821 while PNBP has a negative result and is categorized very low with facilities where the correlation coefficient is -0.131.
2. The results of this study showed that the coefficient of determination of 0.856 or equal to 85.6 %, this figure indicates that the variable Number of Students (X1) and Facility Variables (X2) simultaneously (together) affect the variable PNBP (Y) by 85.6%.
3. From hypothetical test results through F(simultaneous) test and T test (partial). for test F where the value of F count 11.865 > 5.79 significant at 5 % or 0.05, then following the basis of decision making in test F can be concluded that the hypothesis can be accepted with this number of students and facilities simultaneously affect the PNBP and for the test t the number of students partially affects PNBP while the facility does not partially affect PNBP.
4. Based on the results of research multiple linear regression algorithms can predict income in the coming year, from the results of the regression equation obtained that Non-Tax State Revenue in the period 2021 increased compared to the previous year.

5. ACKNOWLEDGMENTS

The authors would like to express their deepest gratitude to all those who have assisted in this study. Especially to the State Polytechnic of Nusa Utara because it has been willing to be a source of data for this research.

6. REFERENCES

[1] Undang-Undang Nomor 9 Tahun 2018 tentang

Penerimaan Negara Bukan Pajak

[2] Peraturan Pemerintah Nomor 22 Tahun 1997 tentang Jenis Dan Penyetoran Penerimaan Negara Bukan Pajak

[3] Nursanti, Masdar Mas'ud, Nur Alam, 2019. Efektivitas dan Pengelolaan Penerimaan Negara Bukan Pajak. PARADOKS,Vol.2, No.4

[4] Ibeto, I., and Justine, C., 2012. Issues And Challenges In Local Government Project Monitoring And Evaluation In Nigeria: The Way Forward. European Scientific Journal, Vol. 8, No.18, ISSN: 1857 – 7881.

[5] Gaspersz,Vincent.2005.Production Planning and Inventory Control.Jakarta:Gramedia Pustaka Utama.

[6] Kutner, M.H., Nachtsheim, C.J., dan Neter, J. (2004). Applied Linear Regression Models. Fourth Edition. McGraw-Hill Companies, Inc., New York.

[7] Heizer, Jay and Render Barry, (2015), Manajemen Operasi : Manajemen Keberlangsungan dan Rantai Pasokan, Salemba Empat, Jakarta.

[8] Sugiyono. (2018). Metode Penelitian Kuantitatif.Bandung: Alfabeta.

[9] Deni Lukman Hakim, Lis Utari, 2020. Prediksi Jumlah Pembelian Sepatu Dengan Penerapan Metode Regresi Linear. Teknologi Informasi dan Sains. Vol. 10 No. 2, hlm. 71-80

[10] Alif Al-Fadhilah Nur Wahyudin , Aji Primajaya , Agung Susilo Yuda Irawan, 2020. Penerapan Algoritma Regresi Linear Berganda Pada Estimasi Penjualan Mobil Astra Isuzu. Techno.COM, Vol. 19, No. 4, hlm. 364-374

[11] Purwantoro, 2017. Analisis Prediksi Penerimaan Mahasiswa Baru Dengan Menggunakan Metode Regresi. Rekayasa Informasi, Vol. 6, No.1.

[12] Yuli Triestini , Bambang Nugroho, Rima H.S. Siburian, 2020.Trend PNBP Sektor Kehutanan Provinsi Papua Barat Pasca Implementasi Kebijakan Si-Puhh Online Dan Self Assesment. Cassowary Vol.3 No. 1 hlm 1 – 10.