ISSN 2502-5872 (Print)
ISSN 2622-3627 (Elektronik)

Volume 8 Nomor 1, Februari 2023, page 17 - 28.

Binary Logistics Regression To Predict The Opportunity Of SNMPTN Graduation In Statistics Study Program Of Tanjungpura University

Syifa Khansa Huriyah[1], Neva Satyahadewi[2], Hendra Perdana[3], Ray Tamtama[4]
[1, 2, 3, 4]Tanjungpura University,
Jl. Prof. Dr. H. Hadari Nawawi, Pontianak, Kalimantan Barat.
Email: syifakhansahuriyah@student.untan.ac.id[1], neva.satya@math.untan.ac.id[2], hendra.perdana@math.untan.ac.id[3], raytamtama@gmail.com[4]

ABSTRACT

The National Selection of State University Entrance or *Seleksi Nasional Masuk Perguruan Tinggi Negeri* (SNMPTN) is one of the selections for high school students seeking higher education. The Statistics Study Program as one of the study programs at Tanjungpura University has a capacity of 20 seats in the SNMPTN. This limited capacity causes prospective students to prepare the right strategy in order to be accepted through the SNMPTN. In this study, logistics regression was used to predict the probability of graduation status on the SNMPTN path in the Statistics Study Program of Untan. Binary logistic regression is a statistical analysis technique for representing the relationship between a response variable with two (binary) categories and one or more predictor variables on a continuous or categorical scale. Data for this study were primary data from a questionnaire that received 93 samples. The response variable used is graduation status (Y) through the SNMPTN in Statistics Study Program, Tanjungpura University classified as 1 (passed) and 0 (not passed). Based on the results of the study, it is known that the variables that have a significant effect on graduation status are the status of choice in Statistics Study Program (X_1), national level achievement ownership (X_3), the average value of Mathematics (X_4), the average value of Chemistry (X_6), Biology average score (X_7), Indonesian average score (X8), and English average score (X_9). Meanwhile, provincial level achievement (X_2) and Physics average (X_5) did not have a significant effect on graduation status. The binary logistic regression model obtained has an accuracy error of 15,05% with an accuracy rate of 84,95%, meaning that this model has a good criteria.

Keywords : SNMPTN, Dichotomy, Binary Logistic Regression.

ABSTRAK

Seleksi Nasional Masuk Perguruan Tinggi, atau SNMPTN, merupakan pilihan seleksi bagi siswa sekolah menengah yang mencari pendidikan yang lebih tinggi. Program Studi Statistika merupakan salah satu program studi di Universitas Tanjungpura yang memiliki daya tampung 20 kursi pada SNMPTN. Daya tampung yang terbatas ini menyebabkan calon mahasiswa harus mempersiapkan strategi yang tepat agar dapat diterima melalui jalur SNMPTN. Pada penelitian ini, regresi logistik biner digunakan untuk memprediksi peluang status kelulusan pada jalur SNMPTN di Program Studi Statistika Untan. Regresi Logistik Biner merupakan suatu metode analisis statistika untuk mendeskripsikan hubungan antara variabel respon yang memiliki dua kategori (biner) dengan satu atau lebih variabel prediktor berskala kategori atau kontinu. Data dalam penelitian ini adalah data primer dari kuesioner terhadap 93 responden. Variabel respon yang digunakan yaitu status kelulusan (*Y*) melalui jalur SNMPTN di Program Studi Statistika Untan yang diklasifikasikan 1 (lulus) dan 0 (tidak lulus). Berdasarkan hasil penelitian diketahui bahwa variabel yang berpengaruh

signifikan terhadap status kelulusan yaitu status pilihan pada Program Studi Statistika Untan (X_1), kepemilikan prestasi tingkat nasional (X_3), nilai rata-rata Matematika (X_4), nilai rata-rata Kimia (X_6), nilai rata-rata Biologi (X_7), nilai rata-rata Bahasa Indonesia (X_8), dan nilai rata-rata Bahasa Inggris (X_9). Sedangkan kepemilikan prestasi tingkat provinsi (X_2) dan nilai rata-rata Fisika (X_5) tidak berpengaruh signifikan terhadap status kelulusan. Model regresi logistik biner yang diperoleh memiliki kesalahan akurasi sebesar 15,05% dengan tingkat keakurasian sebesar 84,95%, artinya model ini memiliki kriteria baik.
Kata kunci : SNMPTN, Dikotomi, Regresi Logistik Biner

How to Cite: Huriyah, S. K., Satyahadewi, N., Perdana, H., Tamtama, R. (2023). Binary Logistics Regression To Predict The Opportunity Of SNMPTN Graduation In Statistics Study Program Of Tanjungpura University. *Mathline: Jurnal Matematika dan Pendidikan Matematika*, 8(1), 17 - 28.

DOI: https://doi.org/10.31943/mathline.v8i1.269

PRELIMINARY

The National Selection of State University Entrance is a student selection path to enter various state universities in Indonesia. Unlike the other entry paths, namely the Joint Selection of State University Entrance (*Seleksi Bersama Masuk Perguruan Tinggi Negeri* or SBMPTN) or the independent path that requires taking a written exam first, *Seleksi Nasional Masuk Perguruan Tinggi Negeri* (henceforth SNMPTN) uses report cards, national exam scores and student academic achievements. SNMPTN is a pathway that is in great demand by students, so it has a high level of competition.

Tanjungpura University (Universitas Tanjungpura or Untan) is one of the universities that implements SNMPTN. Currently, there are 9 faculties with 95 study programs at Tanjungpura University. Statistics is one of 9 study programs in the Faculty of Mathematics and Natural Sciences at Untan. The Higher Education Entrance Test Institute or *Lembaga Tes Masuk Perguruan Tinggi* (LTMPT) sets a quota via SNMPTN path for each study program at 40% of the capacity of the study program concerned (LTMPT, 2021). In 2021, the SNMPTN of the Statistics Study Program of Untan would accommodate 20 seats with a total of 180 applicants (kampusaja.com, 2021). This limited capacity encourages prospective students to prepare the right strategies for acceptance through the SNMPTN pathway.

When making predictions, you need a decision support system that can solve a variety of issues related to many criteria and alternatives. This system can assist students in preparing themselves to graduate at new student admissions by understanding the criteria and determining factors. This study uses data mining as a decision support system to predict graduation status in SNMPTN. Data mining is a set of processes aimed at finding

value from a set of data in the form of unknown knowledge. In data mining, there are many methods used to solve problems such as description, classification, clustering, association, prediction, and estimation (Susanto and Meiryani, 2019). One of the models used in the estimation is logistics regression.

In this study, logistics regression was used to predict the probability of graduation status on the SNMPTN path in the Statistics Study Program of Untan. Logistic regression is a technique that can be used to find relationships between categorical response variables and predictor variables. Binary logistic regression analysis is a logistic regression where the response variable has two categories for example 0 and 1, passed and not passed, male and female and other categories.

METHODS

1. Binary Logistics Regression

Logistic regression is a statistical analysis technique for describing the relationship between a response variable with two or more categories and one or more predictor variables on a categorical or continuous scale (Hosmer, Lemeshow and Sturdivant, 2013). Logistic regression includes several parts such as binary logistic regression, multinomial logistic regression, ordinal logistic regression. This study uses binomial logistic regression.

Binary logistic regression is a data analysis method used to analyze the relationship between several predictor variables and a binary categorical response variable. The response variable solely consists of two categories with $Y = 1$ stating the results obtained are "passed" and $Y = 0$ with "not passed" (Pratiwi and Dewi, 2021). The logistic regression model with predictor variables p is denoted as follows (Safitri, Sudarmin and Nusrang, 2019).

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}} \qquad (1)$$

where,

$\pi(x)$: probability of an event (passed) with the value $0 \leq \pi(x) \leq 1$

β_0 : intercept (constant number)

$\beta_1, \ldots, \beta_p$: logistic regression parameters

$x_1, \ldots, x_p$: predictor variable with a number of p

2. Parameter Estimation

Maximum Likelihood Estimation (MLE) is used for parameter estimation in logistic regression, and basically maximum likelihood estimation produces an estimate by maximizing the likelihood function. Mathematically, the likelihood function of the logistic regression model can be expressed as follows (Peeters, Dewil, and Smets, 2012):

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi(\boldsymbol{x_i})^{y_i} [1 - \pi(\boldsymbol{x_i})]^{1-y_i} \quad (2)$$

Observations on the *i* variable are denoted as y_i, while $\pi(\boldsymbol{x_i})$ is an opportunity for the predictor variable *i*. To make the likelihood function easier to calculate, the likelihood function is maximized in the form of a natural logarithm (Peeters, Dewil, and Smets, 2012) and is expressed as follows:

$$L(\boldsymbol{\beta}) = \ln[l(\beta)]$$
$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i \ln[\pi(\boldsymbol{x_i})] + (1 - y_i)\ln[1 - \pi(\boldsymbol{x_i})]) \quad (3)$$

Logistic regression coefficients $\hat{\beta}$ are estimated by solving the first derivative of the natural logarithm of the likelihood function, $L(\boldsymbol{\beta})$ against β in which the result is equalized to 0.

3. Simultaneous Hypothesis Test

simultaneous hypothesis testing was used to determine the effect of predictor variables simultaneously or at the same time on the response variable. The test hypothesis used is as follows (Ramandhani, Sudarno and Safitri, 2017):

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$

H_1 : with minimum one $\beta_j \neq 0, \text{where } j = 1,2,\ldots,p$

The statistics test used was

$$G = -2ln \frac{\left(\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_0}{n}\right)^{n_0}\right)}{\sum_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{(1-y_i)}} \quad (5)$$

where, $n_1 = \sum_{i=1}^{n} y_i$; $n_0 = \sum_{i=1}^{n} (1 - y_i)$; dan $n = n_1 + n_0$.

The number of observations with the value $Y = 0$ is denoted as n_0, while n_1 is the number of observations with the value $Y = 1$. The *G* statistics test is a likelihood ratio test which follows the Chi-Square distribution, so that when a decision is obtained, a comparison is made with the values in table χ^2. The criteria for rejection area are H_0, if the value $G > \chi^2_{(\alpha,p)}$ or *p-value* < α, and the number of predictor variables is denoted *p*. If H_0 is rejected, it means the predictor variable simultaneously affects the response variable.

4. Partial Hypothesis Test

This test was used to determine the effect of each individually or partially on the response variable. The test hypothesis used for each predictor variable (Ramandhani, Sudarno and Safitri, 2017) was namely:

$H_0 : \beta_j = 0$, dimana $j = 1,2, \dots, p$

$H_1 : \beta_j \neq 0$, dimana $j = 1,2, \dots, p$

The statistics test used in the partial parameter estimation was:

$$W_j = \left[\frac{\beta_j}{SE(\beta_j)}\right]^2, \; where \; SE(\beta_j) = \sqrt{\frac{\pi_j(1-\pi_j)}{n}} \qquad (6)$$

The *W* statistics test is also known as the Wald test. This test follows the Chi-Square distribution. H_0 is rejected if the value of $W > \chi^2_{(\alpha,1)}$ or *p-value* < α, this means that the predictor variables have a large impact on the response variable.

5. Model Fit Test

A model suitability test, also called goodness of fit, aims to examine how well the obtained model explains the observed data. Binary logistic regression evaluates model fit using the Hosmer-Lemeshow test of goodness of fit by comparing observed frequencies. The hypothesis test used is as follows (Pratiwi and Dewi, 2021):

H_0: The model used is appropriate (there is no significant difference between the possible predictions of the model and the observations results).

H_1: The model used is not appropriate (there is a significant difference between the possible predictions of the model and the observed results).

The statistics test used in the model suitability test is:

$$\hat{C} = \sum_{k=1}^{g} \frac{(O_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1-\bar{\pi}_k)} \qquad (7)$$

where:

O_k : Observation in groups-k($\sum_{j=1}^{c_k} y_j$ with c_k : response (0,1))

$\bar{\pi}_k$: Average estimated probability $\left(\sum_{j=1}^{c_k} \frac{m_j \hat{\pi}(x_j)}{n'_k}\right)$

g : Number of groups (category combination in concurrent model)

n'_k : Number of observations in group *k*

m_j : the number of subjects in the c_k predictor combination.

Critical area of rejection of Ho if $\hat{C} > \chi^2_{(\alpha,p-2)}$ or *p-value* < α, and the number of predictor variables is denoted *p*.

6. Parameter Coefficient Interpretation

One of the measures used to interpret the parameter coefficients of the predictor variables is called the odds ratio. Odds ratio is the ratio of the probability of an event occurring to the probability that the event will not occur. Odds ratio is defined as the tendency of the response variable to have a certain value if it is given and compared to. The following equation was used to determine the odds ratio (Pratiwi and Dewi, 2021).

$$OR = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1} \quad (8)$$

A value of 1 for OR indicates that there is no relationship between the two variables. if the value of OR is less than 1, there is a negative relationship between the two variables to the change in the category of the predictor variable. Vice versa, if the value of OR is greater than 1, there is a positive relationship between the two variables to changes in the category of predictor variables.

7. Classification Accuracy

The classification procedure aims to use the apparent error rate (APER) to determine if the data classification is correct. The APER value indicates the percentage of samples incorrectly classified by the classification function (Johnson and Winchern, 2013). Table 1 is a cross tabulation between the results of observations with the estimated classification of the two classes.

Table 1. Confucian Matrix For Two Classes

Actual	*Predict*	
	$Y = 1$	$Y = 0$
$Y = 1$	n_{11}	n_{10}
$Y = 0$	n_{01}	n_{00}

Information:
n_{11} : The number of response variables (Y=1), has been properly classified as (Y=1)
n_{01} : The number of response variables (Y=0), not yet properly classified as (Y=1)
n_{10} : The number of response variables (Y=1), not yet properly classified as (Y=0)
n_{00} : The number of response variables (Y=0), has been properly classified as (Y=1)

The APER value in the calculation results is a proportion that is predicted to be incorrect with the following classification (Ramandhani, Sudarno and Safitri, 2017):

$$APER(\%) = \frac{n_{01} + n_{10}}{n_{11} + n_{01} + n_{10} + n_{00}}$$

While the accuracy or value of classification accuracy is obtained by 100-APER(%).

RESULTS AND DISCUSSION

1. **Descriptive Data**

The data used in this study are primary data obtained from questionnaire results. Students of the Statistics Study Program of Untan classes 2017 to 2021 are the objects of this study. Based on the Academic and Student Administration Bureau Untan, the population of SNMPTN enthusiasts in the Statistics Study Program of Untan was in 2017 to 2021 with a number of 664 students. Non-probability Sampling is the sampling technique used in this study. The number of samples collected after pre-processing the data was 93 respondents consisting of nine predictor variables and one response variable. The response variable used was graduation status (*Y*) through the SNMPTN path in Statistics Study Program Untan classified as 1 (passed) and 0 (not passed). Then, the predictor variables used, including:

- Status of choice in Statistics Study Program Untan (X_1) with an ordinal scale: First Choice (1), Second Choice (2);
- Ownership of achievements at the provincial level (X_2) with a nominal scale: Having achievements (1), Not having achievements (2);
- Ownership of achievement at the national level (X_3) with a nominal scale: Having achievements (1), Not having achievements (2);
- Average grades of report cards from grade 10 to grade XII for several subjects, including Mathematics (X_4), Physics (X_5), Chemistry (X_6), Biology (X_7), Indonesian (X_8), and English (X_9).

Based on the sample data as many as 93 respondents, it was obtained the information that there were 56 respondents who passed the SNMPTN, and the remaining 37 respondents did not pass. The descriptive qualitative data is presented in Table 2, while quantitative data is presented in Table 3.

In Table 2, the obtained information is that there are 76 respondents or 81,72% who chose Statistics Study Program as the first choice. Then, only 4 respondents or 4,30% have achievements at the national level, while respondents who have achievements at the provincial level are 15 respondents or 16,13%.

Table 2. Descriptive Qualitative Variable Data

Variable	Information of Variable	Classification	Total	Percentage
X_1	Choosing Statistics Study Program of Untan	1. First Choice 2. Second Choice	76 17	81,72% 18,28%
X_2	Ownership of achievement at the national level	1. Yes (I have) 2. No	15 78	16,13% 83,87%
X_3	Ownership of achievement at the national level	1. Yes (I have) 2. No	4 89	4,30% 95,70%

In Table 3, it is seen that the average value of report cards from grade 10 to grade 12 for Mathematics is 86,41. Of the six subjects used, it was found that the average value of Physics had the lowest average compared to other subjects. Meanwhile, the highest average is Mathematics.

Table 3. Descriptive Data of the Quantitative Variable

Variable	Information of Variable	Minimum	Average	Maximum
X_4	Math Average Score	75,60	86,41	96,89
X_5	Physics Average Score	74,20	83,71	95,40
X_6	Chemistry Average Score	74,20	83,88	95,80
X_7	Biology Average Score	76,60	84,70	92,80
X_8	Indonesian Average Score	76,00	85,71	91,80
X_9	English Average Score	71,40	84,60	94,60

2. The Results of Analysis

Simultaneous test of the logistic regression model aims to obtain a good and simple model based on factors that can affect the response variable. Based on Table 4, the deviance value (*G*) for the regression model formed is 54,4651 which is larger than the value $\chi_{(9;0,05)}$ = 16,919 so that H_0 is rejected. This means that at least one predictor variable has a large impact on the response variable.

Table 4. Simultaneous Logistic Regression Test

G	*Df*	$\chi^2_{(0,05;9)}$	Decision
54,4651	9	16,919	Ho rejected

Partial tests were then performed to determine the effect of each predictor variable individually or partially on the response variable. Based on Table 5, the variables that has the most impact on graduation status are preferred status in the Statistics Study Program of Untan (X_1), national level achievement ownership (X_3), Mathematics average value (X_4), Chemistry average value (X_6), the average value of Biology (X_7), the average value of Indonesian (X_8), and the average value of English (X_9). This is because the p-value is below the 5% significance level, leading to the decision to reject Ho. Meanwhile, provincial level achievement (X_2) and Physics average (X_5) have no significant effect on graduation status.

Table 5. Logistics Regression Partial Test

Variable	**Estimation (B)**	***exp(B)***	***p-value***
Intercept	5,7704		0,1747
$X_1(2)$	-6,0182	0,0024	0,0000
$X_2(2)$	-1,7492	0,1739	0,0096
$X_3(2)$	0,3013	1,3516	0,4094
X_4	0,1124	1,1190	0,0072
X_5	-0,0636	0,9384	0,1956
X_6	0,3086	1,3615	0,0000
X_7	0,1474	1,1588	0,0000
X_8	-0,2886	0,7493	0,0000
X_9	-0,2462	0,7818	0,0000

Based on Table 5, odds ratio values for all predictor variables listed in column *exp(B)*. In the study program choice status variable (X_1), if a student chooses the Statistics Study Program Untan as the second choice, then the chance to pass is 0,0024 times smaller than the first choice.

In the variable of the average value of the Mathematic report card (X_4), if the student's average score is higher, the chances of passing the SNMPTN in the Statistics Study Program Untan are 1,1190 times greater. Meanwhile, on the variable average value of English report cards (X_9), if the average score of students is higher, the chances of passing the SNMPTN in the Statistics Department of Untan are decreasing. This is because the estimate obtained is negative so that the odds ratio value is < 1.

3. Conformity Test Results of the Binary Logistics Regression Model

The Goodness-of-Fit test was used to evaluate whether the resulting model is feasible or not. Based on Table 6, the resulting chi-square value is 13,086 or less than (0,05;7) = 14,0671. Therefore Ho is not rejected, meaning that the model is suitable or there is no significant difference between the possible predictions of the model and the observations.

Table 6. *Goodness-of-Fit Test*

chi-square	***Df***	$\chi^2_{(0,05;7)}$	**Decision**
13,086	7	14,0671	Ho accepted

4. Interpretation of the Model

The interpretation of the formed logistic regression model is based on Table 5 and uses Equation (1) as follows:

$$\pi(x) = \frac{e^{\left(\begin{matrix}5,7704-6,0182x_{12}-1,7492x_{22}+0,3013x_{32}+0,1124x_4+\\0,0636x_5+0,3086x_6+0,1474x_7-0,2886x_8-0,2462x_9\end{matrix}\right)}}{1+e^{\left(\begin{matrix}5,7704-6,0182x_1-1,7492x_2+0,3013x_3+0,1124x_4+\\0,0636x_5+0,3086x_6+0,1474x_7-0,2886x_8-0,2462x_9\end{matrix}\right)}}$$

An example is given here, student A chooses the Statistics Study Program in the first choice. She does'nt have any provincial or national achievements, then the mean report card score for all subjects is 80. Based on these characteristics, the opportunity values are obtained as follows:

$$\pi(x) = \frac{e^{\left(\begin{matrix}5,7704-6,0182(0)-1,7492(1)+0,3013(1)+0,1124(80)-\\0,0636(80)+0,3086(80)+0,1474(80)-0,2886(80)-0,2462(80)\end{matrix}\right)}}{1+e^{\left(\begin{matrix}5,7704-6,0182(0)-1,7492(1)+0,3013(1)+0,1124(80)-\\0,0636(80)+0,3086(80)+0,1474(80)-0,2886(80)-0,2462(80)\end{matrix}\right)}} = 87,24\%$$

Based on calculations using Equation (1) and the parameter coefficient information in Table 4, it is found that student A who has the opportunity to pass as a student of Statistics Study Program of Untan is as much as 87,24%. Other illustrative examples are presented in Table 7.

Table 7. Illustration of the Calculation of Opportunities for Passing SNMPTN in Statistics Study Program of Untan

Variable	**Student A**	**Student B**	**Student C**	**Student D**
Study program choice	1	1	2	2
National Achievement	None	Yes	None	Yes
Provincial Achievement	None	None	None	None
Mathematics	80	80	80	80

Variable	Student A	Student B	Student C	Student D
Physics	80	80	80	80
Chemistry	80	80	80	80
Biology	80	80	80	80
Indonesian	80	80	80	80
English	80	80	80	80
Chance to pass	**87,24%**	**97,52%**	**1,64%**	**8,73%**

Student B has almost the same characteristics as student A, where the difference is the ownership of national level achievements. Based on the calculations in Table 7, the probability of student B is higher than student A, which is 97,52% compared to 87,24%. Furthermore, student C has different characteristics with student A, namely choosing Statistics Study Program of Untan as the second choice. The calculation results for student C have a chance of passing only 1,64%.

Table 8. Confusion Matrix Results

Actual	Predict	
	$Y = 1$	$Y = 0$
$Y = 1$	50	8
$Y = 0$	6	29

$$APER(\%) = \frac{6+8}{50+6+8+29} = 15{,}05\%$$

$$Accuration\ (\%) = 1 - 0{,}1505 = 84{,}95\%$$

5. Result of Classification Accuracy

Table 8 is the result of the Confusion Matrix calculation for the two classes (binary). Based on the APER calculation, it was found that the value of the classification error was 15,05%. Meanwhile, the percentage of accuracy or classification accuracy is 84,95%. Therefore, the process of determining the graduation status of students on the SNMPTN path by choosing the Statistics Study Program of Untan is in the good category.

CONCLUSION

In a nutshell, the variables that greatly affects on graduation status are the status of choice in the Statistics Study Program of Untan (X_1), national achievement ownership (X_3), Mathematics average score (X_4), Chemistry average score (X_6), Biology average score (X_7), Indonesian average score (X_8), and English average score (X_9). Meanwhile,

provincial level achievement (X_2) and Physics average score (X_5) do not have a significant effect on graduation status. The binary logistic regression model obtained has an accuracy rate of 84,95%, meaning that this model has a good criteria.

From the conclusions above, the suggestions that the authors can convey are: (1) Looking for other variables that affect the acceptance of SNMPTN such as school accreditation and the performance of alumni who are in the destination study program. Hence, it can determine the level of accuracy of the binary logistic regression model when the variables are added. (2) This model was developed into an application, so as to facilitate the guidance and counseling teachers to provide advice to students in determining study programs during SNMPTN. (3) Expanding the research sample not only at the study program level, but also in the scope of the university.

ACKNOWLEDGMENTS

The authors would like to thank FMIPA Tanjungpura University for providing the 2021 DIPA FMIPA Untan research funding in conducting this research study..

REFERENCE

Hosmer Jr., D.W., Lemeshow, S. and Sturdivant, R.X. (2013). *Applied Logistic Regression 3nd*. John Wiley and Sons, Hoboken, NJ.

Johnson, R.A., and Winchern, D.W. (2013). *Applied Multivariate Statistical Analysis 6th*. Prentice Hall New Jersey.

kampusaja.com. (2021). Daya Tampung SNMPTN Untan. Diakses 5 November 2021, dari https://kampusaja.com/daya-tampung-snmptn-untan/

LTMPT. (2021), Jalur Penerimaan dan Alokasi Daya Tampung Mahasiswa Baru. Diakses 5 November 2021, dari https://www.ltmpt.ac.id/?mid=10

Peeters, B., Dewil, R., and Smets, I.Y. (2012). Improved Process Control of an Industrial Sludge Cemtrifuge-driver Installation Through Binary Logistic Regression Modeling of The Fouling Issues. *Journal of Process Control, Vol 22: 1390-1391*.

Pratiwi, R., and Dewi, A.F. (2021). Analisis Regresi Logistik Biner pada Pengaruh Harga, Kualitas Pelayanan dan Promosi terhadap Kepuasan Pelanggan dalam Menggunakan Jasa Layanan Grab di Kabupaten Lamongan. *Inferensi*, vol. 4, no. 2, p. 77.

Ramandhani, R., Sudarno, and Safitri, D. (2017). Metode Bootstrap Aggregating Regresi Logistik Biner untuk Ketepatan Klasifikasi Kesejahteraan Rumah Tangga di Kota Pati. *Gaussian, Vol 6, No 1: 121-130.*

Safitri, A., Sudarmin, and Nusrang, M. (2019). Model Regresi Logistik Biner pada Tingkat Pengangguran Terbuka di Provinsi Sulawesi Barat Tahun 2017. *VARIANSI J. Stat. Its Appl. Teach. Res., vol. 1, no. 2, p. 1.*

Susanto, A., and Meiryani. (2019). Functions, processes, stages and application of data mining. *Int. J. Sci. Technol. Res., vol. 8, no. 7, pp. 136–140.*