Mathematical Journal of Modelling and Forecasting
Vol. 1, No. 2, December 2023, pp. 1-9

University Election Analysis: Logistic Regression Approach with Dummy and Ordinal Variables

Miftha Delinda[1] and Devni Prima Sari[1,*]

[1] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Padang, Indonesia
[*]devniprimasari@fmipa.unp.ac.id

Abstract. Education has a significant role in advancing the development of a country. One of them is the university. Thus, if continuing studies at university, it is hoped that the knowledge and skills following the study program will later become the primary capital to be more competent in the world of work. Logistic regression is a statistical method that can be used to determine the factors that influence the choice of university for class XII Phase F students. The dependent variable consists of two categories. This research aims to determine the factors that influence the choice of university for class XII Phase F students. This type of research is applied research and uses primary data obtained from filling out questionnaires. This research was carried out at SMAN 3 Padang. The population in this study were all students in class XII Phase F at SMA Negeri 3 Padang, with a sample of 78 students obtained using the Purposive Sampling method. The results of the research show that the factors that influence the choice of university for class XII Phase F students at SMA Negeri 3 Padang are father's work based on educated and trained labor, father's work based on educated labor, father's work based on trained labor, father's work based on uneducated and unskilled labor, convenient university location, easy access to transportation, rent, food and daily living costs-affordable days according to budget, and information from social media and websites.

Keywords: Logistic Regression, University, High School

1. Introduction

Education has an essential role in developing a country with high work productivity. This involves planned efforts to develop one's potential to become a human being with intelligence, character, and competence, as well as a complete human being [1]. Higher education prepares competitive, reliable human resources [2]. The next level of education after high school is college, which includes various programs such as diploma, bachelor's, master's, specialist, and doctorate. In Indonesia, higher education institutions include polytechnics, colleges, and universities [3].

Thus, if continuing studies at university, it is hoped that the knowledge and skills the study program is taking will be the essential capital to be more competent in the world of work [4]. Considering that competition in the world of work is increasingly fierce, many job opportunities require workers with a diploma or bachelor's degree. Therefore, taking secondary education alone is not enough to compete in the world of work to obtain job opportunities [5].

Many factors influence a student's choice of university. According to [6], factors influencing students' choice of university include reference groups, family, university reputation, personal factors, university location, job prospects, university costs, financial aid/scholarships, and information sources. Research [7] found that six factors influenced SMA 22 Surabaya students' choice of university: reference group, institutional image, motivation, family, location, and price. Based on [8], the reference group is one or more individuals who are used as comparisons to form emotional responses and describe a person's behavior. Meanwhile, in research conducted by [9] using multinomial logistic regression, the results showed that five independent variables influenced the choice of college for class XII students in Bojonegoro Regency, namely major, number of dependents, father's job, father's education, and mother's education.

Based on the description above, researchers want to know what factors influence the choice of university for class XII Phase F students, so an analysis is needed. One analysis that can describe a cause-and-effect relationship is regression analysis. If the dependent variable is categorical, logistic regression analysis is used [10]. The logistic regression method is a regression approach that is useful for analyzing the relationship between a dependent variable (Y), which has categorical characteristics, and one or more independent variables (X), which have continuous, categorical, or both characteristics [11]. Based on [12], the logistic regression model is the most commonly used to analyze data where the response variable has categorical characteristics.

2. Regression Analysis

Regression analysis is an analysis that aims to determine the effect of a variable on other variables. The simplest regression model is a simple linear regression model with the equation form [13]:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

where:

$Y =$ dependent variable
$X =$ independent variable
β_0 = constant
β_1 = regression coefficient
ε = random error

2.1 Ordinal Scale

The ordinal scale shows an order (rank or level or ranking), which also functions as a grouping. For example, education level is categorized as "0" for elementary school, "1" for middle school, and "2" for high school [14].

2.2 Dummy Variable

Dummy variables are used to quantify qualitative variables (for example, gender, race, religion, changes in government policy, differences in situations, etc.). The following are the conditions for forming dummy variables [15]:

a. Variable with k categories will require $k - 1$ dummy variables.
b. The dummy variable is binary. The values are 0 and 1, where the value 1 indicates the presence of the variable while 0 indicates the absence of the variable.

2.3 Logistic Regression Analysis

Logistic regression analysis is a regression method that can be used to describe the relationship between a dependent variable (Y), which is categorical, and one or more independent variables (X) which are continuous, categorical, or both [11]. The form of the logistic regression model is [16]:

$$\pi(x_i) = \frac{e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p+\varepsilon)}}{1 + e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p+\varepsilon)}} \quad (2)$$

To make it easier to estimate regression parameters, $\pi(x_i)$ in the equation above is transformed to produce the logit form of logistic regression, as follows:

$$g(x) = \ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \quad (3)$$

2.4 Parameter Estimation

The method used to estimate logistic regression parameters is MLE (Maximum Likelihood Estimation). The Maximum Likelihood Estimation method provides an estimated value of β to maximize the likelihood function [17]. Systematically, the likelihood function for the logistic regression model is as follows [16]:

$$l(\beta) = \prod_{i=1}^{n} \pi(x_i)^{y_i}(1 - \pi(x_i)^{1-y_1} \tag{4}$$

where:

y_i = observation on the ith variable
$\pi(x_i)$ = probability for the ith independent variable

2.5 Test the Logistic Regression Model

The model test was carried out to examine the role of the independent variables in the model together. The simultaneous test is also called the chi-square model test [17]. The hypothesis in this test is

$$H_0 : \beta_0 = \beta_1 = \cdots = \beta_k = 0$$
$$H_1 : \text{there is } \beta_j \neq 0 \; ; j = 1,2, \ldots, k$$

with test statistics:

$$G = -2 \ln \left[\frac{\text{likehood without the variable}}{\text{likehood with the variable}} \right]$$
$$G = -2 \ln \left[\frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^{n} \hat{\pi}_i{}^{y_i}(1 - \hat{\pi}_i)^{1-y_i}} \right] \tag{5}$$

where:
n_0 = how many y_i have the value 0
n_1 = how many y_i have the value 1
n = many y_i $(n_0 + n_1)$

The G-test statistic follows a chi-square distribution (X^2) with k degrees of freedom. With the test criteria $G > X^2{}_{(k,\alpha)}$, reject H_0, which means that in the regression model, at least one parameter is not equal to zero. In other words, this model may be recommended, but it is not the best model, and the analysis can be continued by looking for the best model [16].

2.6 Partial Hypothesis Testing

A partial test is carried out to determine whether the independent variable significantly affects the dependent variable alone or not. Partial or individual test results will show whether an independent variable is suitable to be included in the model [16].
Hypothesis:

$H_0 : \beta_j = 0$ (variable x_j has no real effect)
H_1 : there is $\beta_j \neq 0$, (variable x_j has a significant effect) for $j = 1,2, \ldots, k$

The general formula for Wald test statistics for logistic regression tests according to [13].

$$wald\ (W_j) = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}, j = 1,2, \ldots, k \tag{6}$$

with $SE(\hat{\beta}_j) = [var(\hat{\beta}_j)]^{1/2}$
where:

$\hat{\beta}_j$ = parameter estimator
$SE(\hat{\beta}_j)$ = standard error of the parameter estimator

The ratio resulting from the test statistic, under the hypothesis H_0, will follow the standard normal distribution. So, to obtain a decision, a comparison is made with the standard normal distribution (Z). Rejection criteria (reject H_0) if $W > Z_{\alpha/2}$ or the significance value is less than α, then reject H_0.

2.7 Model Backward Elimination

Backward elimination is a step backward by regressing all independent variables with the dependent variable and eliminating the independent variables based on the most significant p-value. Backward elimination is a suitable regression method because the behavior of the independent variables is explained as well as possible by selecting the independent variables from the many independent variables available in the data.

2.8 Coefficient Interpretation

Interpretation of the coefficients in the logistic regression model is carried out in the form of odds ratios, which aims to see the extent of the real influence of the independent variables. The odds ratio is the risk or tendency to experience a specific event between one category and another [18]. The odds ratio value is written as follows [16]:

$$\psi = \frac{\frac{\pi(1)}{[1-\pi(1)]}}{\frac{\pi(0)}{[1-\pi(0)]}} = \frac{e^{(\beta_o+\beta_1)}}{e^{(\beta_o)}} = e^{\beta_1} \quad (7)$$

If the value $\psi = 1$, then there is no relationship between the two variables. If the value $\psi < 1$, then between the two variables, there is a negative relationship with the change in category of the x value and vice versa if $\psi > 1$.

3. Research Methods

This research is applied to class XII Phase F students of SMA Negeri 3 Padang for the 2023/2024 academic year. The data type used in this research is primary data obtained by distributing questionnaires [19]. The total sample in the study was 78 students. This research uses a Non-Probability Sampling technique with the Purposive Sampling method.

The variables used in this research consist of the dependent variable, namely students who choose universities in West Sumatra and students who choose universities outside West Sumatra as well as the independent variable, namely:

a. Father's job (X_1) is on a nominal scale
   Father's work has five categories: father's work is based on educated and skilled labor, father's work is based on educated labor, father's work is based on trained labor, father's work is based on uneducated and unskilled labor, and father not working. Thus, a dummy variable was formed with a reference or base category, namely, father not working, as follows:
   $Dummy\ variable = k - 1 = 5 - 1 = 4$
   The values given are:
   D_{11} = 1 if the father's work is based on educated and skilled labor, =0 otherwise
   D_{12} = 1 if the father's work is based on educated labor, =0 otherwise
   D_{13} = 1 if the father's work is based on trained labor, =0 otherwise
   D_{14} = 1 if the father's work is based on uneducated and unskilled labor, =0 otherwise
b. Mother's job (X_2) is on a nominal scale
   Mother's work has five categories: mother's work is based on educated and skilled labor, mother's work is based on educated labor, mother's work is based on trained labor, mother's work is based on uneducated and unskilled labor, and mother not working. So, a dummy variable was formed with a reference or base category, namely, mother not working, as follows:
   $Dummy\ variable = k - 1$
   $= 5 - 1$
   $= 4$
   The values given are:

D_{11} = 1 if mother's work is based on educated and skilled labor, =0 otherwise
D_{12} = 1 if mother's work is based on educated labor, =0 otherwise
D_{13} = 1 if mother's work is based on trained labor, =0 otherwise
D_{14} = 1 if mother's work is based on uneducated and unskilled labor, =0 otherwise

c. Father's income (X_3) is on an ordinal scale
d. Mother's income (X_4) is on an ordinal scale
e. The university's accreditation (X_5), likert scale.
f. The study programs offered are by student interests (X_6), likert scale.
g. Number of job opportunities (X_7), likert scale.
h. Scholarships offered (X_8), likert scale.
i. Convenient university location (X_9), likert scale.
j. Easy transportation access (X_{10}), likert scale.
k. Rent, food and daily necessities are affordable according to the budget (X_{11}), likert scale.
l. Parental recommendation (X_{12}) likert scale.
m. Guidance teacher recommendations (X_{13}), likert scale.
n. Peer influence (X_{14}), likert scale.
o. There were school visits by school representatives (X_{15}), likert scale.
p. Information from social media and websites (X_{16}), likert scale.
q. Alumni experiences and success stories (X_{17}), likert scale.

4. Results and Discussion

The steps that must be carried out in logistic regression analysis are as follows:

a. Estimation of logistic regression model parameters.

The estimation of model parameters was carried out using MLE (Maximum Likelihood Estimator), the likelihood function values by including the 23 independent variables resulting from the estimated parameters of the logistic regression model with the maximum likelihood shown in Table 1.

Table 1. Results of Estimated Logistic Regression Parameters with All Independent Variables

Independent Variable	B
D_{11}	4.072
D_{12}	6.063
D_{13}	5.320
D_{14}	5.067
D_{21}	-1.207
D_{22}	-0,478
D_{23}	-0.392
D_{24}	-0,725
X_3	-0,431
X_4	0.562
X_5	0.925
X_6	-0.609
X_7	-0.447
X_8	-0.199
X_9	-1.697
X_{10}	2.323
X_{11}	2.275
X_{12}	-0.252
X_{13}	-0.160
X_{14}	-0.437
X_{15}	1.308
X_{16}	-1.548
X_{17}	-0.166
Constant	-8.595

b. Logistic regression estimation model

Based on Table 1 above, the logistic regression conjecture model referring to equation (2) involving all independent variables can be obtained as follows:

$$\pi(x_i) = \frac{e^{\begin{pmatrix} -8.595+4.072D_{11}+6.063D_{12}+5.320D_{13}+5.067D_{14}-1.207D_{21}-0.478D_{22}-0.392D_{23} \\ -0.725D_{24}-0.431x_3+0.562x_4+0.925x_5-0.609x_6-0.447x_7-0.199x_8-1.697x_9+2.323x_{10} \\ +2.275x_{11}-0.252x_{12}-0.160x_{13}-0.437x_{14}+1.308x_{15}-1.548x_{16}-0.166x_{17} \end{pmatrix}}}{1+e^{\begin{pmatrix} -8.595+4.072D_{11}+6.063D_{12}+5.320D_{13}+5.067D_{14}-1.207D_{21}-0.478D_{22}-0.392D_{23} \\ -0.725D_{24}-0.431x_3+0.562x_4+0.925x_5-0.609x_6-0.447x_7-0.199x_8-1.697x_9+2.323x_{10} \\ +2.275x_{11}-0.252x_{12}-0.160x_{13}-0.437x_{14}+1.308x_{15}-1.548x_{16}-0.166x_{17} \end{pmatrix}}}$$

By carrying out a transformation, namely the logit transformation of $\pi(x_i)$ above, it can fulfill linear properties. So, the estimated model produces the following logit form:

$$\begin{aligned} g(x_i) = -8.595 &+ 4.072D_{11} + 6.063D_{12} + 5.320D_{13} + 5.067D_{14} - 1.207D_{21} - 0.478D_{22} - 0.392D_{23} \\ &- 0.725D_{24} - 0.431x_3 + 0.562x_4 + 0.925x_5 - 0.609x_6 - 0.447x_7 - 0.199x_8 \\ &- 1.697x_9 + 2.323x_{10} + 2.275x_{11} - 0.252x_{12} - 0.160x_{13} - 0.437x_{14} + 1.308x_{15} \\ &- 1.548x_{16} - 0.166x_{17} \end{aligned}$$

c. Testing the significance of the estimated logistic regression model

Then, to test the significance of the model by including all independent variables, the G test was used. The results of the model significance test using the G test can be seen in Table 2.

Table 2. Goodness Test of Estimated Model (Full Model)

Logistic Regression Model	Chi-Square	Sign.
	44.625	.001

Hypothesis testing in this research uses a real level of 0.05. Based on Table 2, the Chi-square value obtained is 44.625 while the $X^2_{(23;0,05)} = 35.17$ value. It can be seen that the value $G > X^2_{(k,\alpha)}$ means that H_0 is rejected. It can also be seen that the significance value of the logistic regression model is smaller than the real level of 0.05, which also concludes that when rejecting H_0, which means that in the regression model, at least one parameter estimator is not equal to zero. In other words, this alleged model can be suggested, but it is not the best model, and analysis can be carried out by looking for the best model by reducing the independent variables from the model. To find out what variables need to be reduced from the model, a Wald test is first carried out to determine the significance of the variables in the model.

d. Testing the significance of logistic regression parameters

Parameter significance testing is used to see the influence of each independent variable on the dependent variable with the help of software, so the results obtained are as in Table 3.

Table 3. Significance Test Results of Estimated Logistic Regression Parameters with All Independent Variable

Independent Variable	B	SE β	Wald	Sign
D_{11}	4.072	2.188	3.463	0.063
D_{12}	6.063	4.147	2.138	0.144
D_{13}	5.320	2.316	5.276	0,022
D_{14}	5.067	2.147	5.572	0,018
D_{21}	-1.207	1.797	0.451	0.502
D_{22}	-0,478	3.467	0.019	0.890
D_{23}	-0.392	2.212	0.031	0.859
D_{24}	-0,725	1.346	0.290	0.590
X_3	-0,431	0.727	0.351	0.554
X_4	0.562	0.983	0.326	0.568
X_5	0.925	0.871	1.127	0.288
X_6	-0.609	0.979	0.386	0.534
X_7	-0.447	0.718	0.387	0.534
X_8	-0.199	0.970	0.042	0.838

Independent Variable	B	SE β	Wald	Sign
X_9	-1.697	0.965	3.091	0.079
X_{10}	2.323	0.863	7.243	0.007
X_{11}	2.275	0.946	5.789	0.016
X_{12}	-0.252	0.838	0.162	0.688
X_{13}	-0.160	0.494	0.036	0.849
X_{14}	-0.437	0.723	0.785	0.376
X_{15}	1.308	0.723	3.274	0.070
X_{16}	-1.548	0.862	3.225	0.073
X_{17}	-0.166	0.613	0.073	0.787
Constant	-8.595	4.744	3.283	0.070

Based on Table 3, of all the independent variables above, it turns out that 19 independent variables have no significant effect on the dependent variable, namely the choice of university for class XII Phase F students, where the parameter significance value is more than 0.05. So, we get a logistic regression model with all independent variables that must be reduced.

e. Selection of the logistic regression model

To select a logistic regression model, use backward elimination, namely simplifying the model by removing one by one the independent variables that have significance greater than the 0.05 real level. It can be seen that the variable that has been reduced with a significance value of less than 0.05 will be the independent variable for the best model that describes the factors that influence the choice of university for class XII Phase F students for the 2023/2024 academic year. Of the 23 independent variables, only eight independent variables were the best model in this research; the father's work is based on educated and trained labor (D_{11}), father's work is based on educated labor (D_{12}), father's work is based on trained labor (D_{13}), father's work is based on uneducated and unskilled labor (D_{14}), convenient university location (x_9), easy access to transportation (x_{10}), rent, food, and daily living costs -affordable days according to budget (x_{11}), and information from social media and websites (x_{16}). To determine the effect of the best model above, it can be seen in the statistical value of the G test based on equation (5) obtained in Table 4.

Table 4. Reduction Model Goodness Test

Logistic Regression Model	Chi-square	Sign.
	38.351	0,001

Based on Table 4, it is known that the statistical value of the Chi-square test from the logistic regression model with reduced independent variables is 38.351, while the value of $X^2_{(8;0,05)} = 15{,}50$. It can be seen that the value of $G > X^2_{(k,\alpha)}$ thus rejects H_0. It also appears that the significance value of the logistic regression model is smaller than the real level of 0.05, which also identifies the rejection of H_0, which means that the reduction model obtained is different from the model with all independent variables. Therefore, the Wald value and significance of the independent variables in the reduced model can be seen in Table 5.

Table 5. Results of logistic regression analysis

Independent Variable	B	SE β	Wald	Sign
D_{11}	3.215	1.639	3.847	0.050
D_{12}	5.184	2.561	4.098	0.043
D_{13}	4.729	1.734	7.440	0.006
D_{14}	4.791	1.654	8.388	0.004
X_9	-1.484	0.747	3.948	0.047
X_{10}	2.110	0.755	7.818	0.005
X_{11}	1.928	0.782	6.079	0.014
X_{16}	-1.604	0.647	6.154	0.013
Constant	-9.789	3.513	7.765	0.005

Based on Table 5, the model was simplified by only linking the independent variables resulting from the reduction to obtain the following:

$$\pi(x_i) = \frac{e^{(-9.789+3.215D_{11}+5.184D_{12}+4.729D_{13}+4.791D_{14}-1.484x_9+2.110x_{10}+1.928x_{11}-1.604x_{16})}}{1+e^{(-9.789+3.215D_{11}+5.184D_{12}+4.729D_{13}+4.791D_{14}-1.484x_9+2.110x_{10}+1.928x_{11}-1.604x_{16})}}$$

with the logit form of $\pi(x_i)$ namely:

$$g(x_i) = -9.789 + 3.215D_{11} + 5.184D_{12} + 4.729D_{13} + 4.791D_{14} - 1.484x_9 + 2.110x_{10} + 1.928x_{11} - 1.604x_{16}$$

f. Interpretation of coefficients

We can look at the odds ratio value to determine how big the tendency is to experience a certain event between one category and another. The odds ratio value can be seen in Table 6.

Table 6. Odds ratio value for logistic regression model

Independent Variable	Exp (β)
D_{11}	24.914
D_{12}	178.336
D_{13}	113.225
D_{14}	120.427
X_9	0.227
X_{10}	8.251
X_{11}	6.873
X_{16}	0.201
Constant	0.000

In interpreting how much influence the independent variables have on university selection for class XII Phase F students, it can be seen from the Odds Ratio value. Based on Table 6, it is known that the variable father's work is based on educated and trained labor (D_{11}) has an odds ratio value of 24.914, father's work is based on educated labor (D_{12}) has an odds ratio value of 178.336, father's work is based on trained labor (D_{13}) has an odds ratio value of 113.225, father's work is based on uneducated and unskilled labor (D_{14}) has an odds ratio value of 120.427, convenient university location (x_9) has an odds ratio value of 0.227, easy access to transportation (x_{10}) has an odds ratio value of 8.251, rent, food and daily living costs -affordable days according to budget (x_{11}) has an odds ratio value of 6.873, and information from social media and websites (x_{16}) has an odds ratio value of 0.201.

5. Conclusion

Based on the results of the research and discussion, the following conclusions can be drawn from the logistic regression model, which describes the factors that influence the choice of university for class XII Phase F students at SMA Negeri 3 Padang for the 2023/2024 academic year, namely:

$$\pi(x_i) = \frac{e^{(-9.789+3.215D_{11}+5.184D_{12}+4.729D_{13}+4.791D_{14}-1.484x_9+2.110x_{10}+1.928x_{11}-1.604x_{16})}}{1+e^{(-9.789+3.215D_{11}+5.184D_{12}+4.729D_{13}+4.791D_{14}-1.484x_9+2.110x_{10}+1.928x_{11}-1.604x_{16})}}$$

The factors that influence the choice of university for class XII phase F students at SMA Negeri 3 Padang are father's work is based on educated and trained labor, father's work is based on educated labor, father's work is based on trained labor, father's work is based on uneducated and unskilled labor, convenient university location, easy access to transportation, rent, food, and daily living costs -affordable days according to budget, and information from social media and websites.

References

[1] N. Munira, "Faktor-Faktor Yang Mempengaruhi Minat Siswa Madrasah Aliyah Negeri (Man) 2 Jembrana Kelas Xii Untuk Melanjutkan Studi Ke Perguruan Tinggi Pada Tahun Ajaran 2017/2018," *J. Pendidik. Ekon. Undiksha*, vol. 9, no. 2.

[2] D. Abdul, L. M. Aliyyuida, and E. Riyanto, *Persepektif Pendidikan Indonesia di Era Globalisasi*, vol. 1, no. 1. 2019.

[3] *Undang-Undang Republik Indonesia tentang Pendidikan Tinggi*. Republik Indonesia, 2012.

[4] R. Ramadhan, Marwan, and Armiati, "Pengaruh Status Sosial Ekonomi Orang Tua dan Motivasi Belajar terhadap Minat Melanjutkan Studi Ke Perguruan Tinggi," *J. Pendidik. Ekon.*, vol. 1, no. 1, pp. 69–84, 2018.

[5] S. Usna. S and Nuraisyiah, "Pengaruh Motivasi Belajar Dan Pendapatan Orang Tua Terhadap Minat Siswa Melanjutkan Studi Ke Perguruan Tinggi Pada Siswa Program Keahlian Akuntansi SMK Negeri 1 Makassar," *J. Pendidik. Akutansi*, 2022.

[6] O. Tamtekin Aydın, "University Choice Process: A Literature Review on Models and Factors Affecting the Process," *Yuksekogretim Derg.*, vol. 5, no. 2, pp. 103–111, 2015,

[7] A. Puspitasari and F. D. Patrikha, "Analisis Faktor-Faktor Yang Mempengaruhi Keputusan Pemilihan Universitas Pada Siswa Kelas XII SMA Negeri 22 Surabaya," *JPEKA J. Pendidik. Ekon. Manaj. dan Keuang.*, vol. 2, no. 1, p. 1, 2018, doi: 10.26740/jpeka.v2n1.p1-10.

[8] Nugroho Setiadi, *Perilaku Konsumen. Cetakan Keempat*, Jakarta: Kencana Prenada Media Group, 2010.

[9] U. F. Kusumawati, *Analisis Faktor-Faktor yang Mempengaruhi Pilihan Perguruan Tinggi dengan Regresi Logistik Multinomial*. 2016.

[10] J. Harlan, *Analisis Regresi Logistik*, Pertama. Depok: Gunadarma, 2018.

[11] A. Agesti, *Categorical Data Analysis*, Second. Florida: A John Wiley & SONS, INC., 2002.

[12] D. G. Kleinbaum and M. Klein, *Logistic Regression*, Third. USA: Springer, 2010.

[13] Suyono. 2015. Analisis Regresi untuk Penelitian. Deepublish. Yogyakarta.

[14] Syahril, *Statistik*, Satu. Padang: Sukabina Press, 2010.

[15] Sembiring. 1995. Analisis Regresi. Bandung: ITB Bandung

[16] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, Second. Canada: A Wiley-Interscience, 2000.

[17] Y. Tampil, H. Komaliq, and Y. Langi, "Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado," *d'CARTESIAN*, vol. 6, no. 2, p. 56, 2017.

[18] M. Z. Mahmudin, A. Rindengan, and W. Weku, "Penggunaan Association Rule Data Mining Untuk Menentukan Pola Lama Studi Mahasiswa F-MIPA UNSRAT," *d'CARTESIAN*, vol. 3, no. 1, p. 1, 2014.

[19] Sugiyono, *Metode penelitian pendidikan : Pendekatan kuantitatif, kualitatif* , 22nd ed. Alfabeta, 2015.