Characteristics of Graduation Predicate UM Buton Students Using Ordinal Logistic Regression

Firman Saputra[1] , Anna Islamiyati[2] , Nirwan[3]
[1,2,3] Department of Statistics, Hasanuddin University, Makassar, 90245, Indonesia
[1] saputraf20h@student.unhas.ac.id, [2] annaislamiyati@unhas.ac.id

Abstract: Predicate graduation students has become power pull by stakeholders, especially existence formation special for graduates with cum laude predicate. Many factors can influence graduate to get predicate his graduation. Based ordinal logistics regression analysis, significant factors influence predicate graduation UM Buton students in 2020-2021 are gender, origins area, number graduate of the study program, tuition payment category and learning methods. Characteristics of probability for women graduates to obtain higher predicate is greater than of men, well as the probability of graduates with online-based learning are better than conventional-based methods.

Keywords— predicate graduation student; logistics regression; ordinal logistics regression; characteristics odds ratio

1. INTRODUCTION

The vision of the Muhammadiyah Buton of University (UM Buton) is become college superior and powerful global competitiveness with destination produce graduates who have competence based maritime. Importance development competence graduate of this has poured in regulation standard graduate of UM Buton students who refer to the regulation government. Every student who passed require to have minimum GPA of 2,00 and academics load at least 144 credits [1].

Graduation student could give predicate for GPA of 2,76-4,00 and GPA of 2,00-2,75 not have predicate. Predicate graduation student divided in three category that is cum laude, very satisfying and satisfying predicate [1]. This predicate has become power pull alone in requirements get work, for example availability formation special for predicate cum laude. The achievement of this predicate could influence from individual internal and external factors. The factors that can influence predicate graduation according to [2] and [3] are gender, origin school, father's occupation, mother's occupation, origin faculty, and origin area. However, from the research not yet see other factors like learning method, cost education payment and number of graduates based on the study program.

View from nature and scale data measurement, the response variable of the student's graduation predicate is categorical with ordinal scale measurement. One the methods used to determine the characteristics of relationship whose response variable with ordinal scale is ordinal logistics regression. Therefore, this research purpose to determine the characteristics of factors that influence the graduation predicate of UM Buton students by using ordinal logistic regression. The results are expected to provide additional information in the formulation of policies related to characteristics graduated at UM Buton.

2. METHODOLOGY

Muhammadiyah Buton of University is one college located in Baubau City, Southeast Sulawesi Province and every year produce thousand graduates. The data used is student graduation in 2020 and 2021 obtained from the database of UM Buton academic system. The suspected factors influence predicate of student graduation (Y) are gender (X_1), origins area (X_2), number graduate of the study program (X_3), category of tuition payment (X_4) and category of learning method (X_5). Number of data collected as many as 2.260 graduates from of thirteen study programs. Process analysis and visualization use STATA application.

2.1 Ordinal Logistics Regression

Ordinal logistic regression is a type of regression that models the relationship between the ordinal scale response variable and the numerical or categorical explanatory variables. The model that can be used for ordinal logistic regression is the logit model contained in the cumulative probability. Suppose that Y is ordinal response variable with category C and $x' = (1, 2, \ldots, x_j)$ is a vector of explanatory variables then the probability c category at value x can be expressed by [3]:

$$P[Y \le c|x] = \pi_c(x) \qquad (1)$$

and function cumulative probability are:

$$P[Y \le c|x] = \pi_1(x) + \cdots + \pi_c(x) \qquad (2)$$

So that the cumulative logit model of ordinal logistics regression defined as:

$$
\begin{aligned}
L_c(x) &= logit(P[Y \le c|x]) \\
&= log\left(\frac{P[Y \le c|x]}{1 - P[Y \le c|x]}\right) \\
&= log\left(\frac{\pi_1(x) + \cdots + \pi_c(x)}{\pi_{c+1}(x) + \cdots + \pi_C(x)}\right) \\
&= \alpha_c + x'\beta
\end{aligned}
\qquad (3)
$$

where $c = 1, \ldots, C - 1$ and $\alpha_c = \alpha_1, \ldots, \alpha_{C-1}$ is a constant vector and β are the coefficient vectors of model.

Method parameter estimation used in this model use method *maximum likelihood* with to do transformation *likelihood* function into the natural logarithm (**ln**) like following [3]:

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{c=1}^{C} y_{ci} \ln[\pi_c(\boldsymbol{x})] \tag{4}$$

where $\boldsymbol{\pi_c(x)}$ is function of parameters $\boldsymbol{\beta}$.

2.2 Parameter Test

The results of parameter estimation from the model that have been obtained need tested the significance of the parameters β, that is simultaneously testing and partial testing.

a) Simultaneous Test

Simultaneous test conducted with destination for knowing parameter β significance from all variable explanatory to the variable response together. The hypotheses in this test as follows:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_j$
H_1: at least there one $\beta_p \neq 0$, with $p = 1, 2, \ldots j$

Testing this hypothesis using the G statistic test or Likelihood Ratio Test (LRTs) as follows [4]:

$$G = -2\ln\left(\frac{L_0}{L_1}\right) \tag{5}$$

where L_0 is the *likelihood* value in the model without explanatory variable and L_1 is *likelihood* value in the model with involve all explanatory variable. The decision rule for the LRTs test is H_0 rejected if $G_{hit} > \chi^2_{\alpha,p}$ or *p-value* $< \alpha$ [4].

b) Partial Test

Partial test was conducted to determine the significance β_j parameter of the explanatory variable by partially. This test hypothesis are:

H_0: $\beta_p = 0$
H_1: $\beta_p \neq 0$, with $p = 1, 2, \ldots j$

Statistics test of this hypothesis is using Wald's test as following [4]:

$$W_p = \left[\frac{\hat{\beta}_p}{\widehat{SE}(\hat{\beta}_p)}\right]^2 \tag{6}$$

where $\hat{\beta}_p$ is the parameter estimator of p, and $\widehat{SE}(\hat{\beta}_p)$ is the standard deviation of $\hat{\beta}_p$. Rule Wald's test decision is H_0 rejected if $W_p > \chi^2_{\alpha,1}$ or *p-value* $< \alpha$ [4].

2.3 Goodness of Fit Model

Testing the goodness of fit model is used to determine whether the model that has been formed good model is statistically. Test statistics used to test goodness of the ordinal logistics regression model is *deviance* test [5].

$$D = -2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1 - y_i)\ln(\frac{1 - \hat{\pi}_i}{1 - y_i})\right] \tag{7}$$

Hypothesis in this testing are:

H_0: model fit on the data
H_1: model doesn't fit on the data

Rule *deviance* test decision is H_0 rejected if $D > \chi^2_{\alpha,(n-p)}$ or *p-value* $< \alpha$ [5].

2.4 Interpretation Parameter

Interpretation of parameters in the ordinal logistics regression model carried out based on value Odds Ratio parameter (comparison risk) with destination for determine connection functional among variable explanatory and variable response as well as explain change difference trend between category from the response variable [5]. Odds Ratio by category $Y \leq c$ defined by:

$$OR_c = \exp(\alpha_c) \tag{8}$$

OR value > 1 indicates the probability of event occurring in the first category is greater than the second category. While the value of OR < 1 indicates that the probability of events occurring in the first category is smaller than the second category [5].

3. RESULT

3.1 Characteristics Graduation Predicate

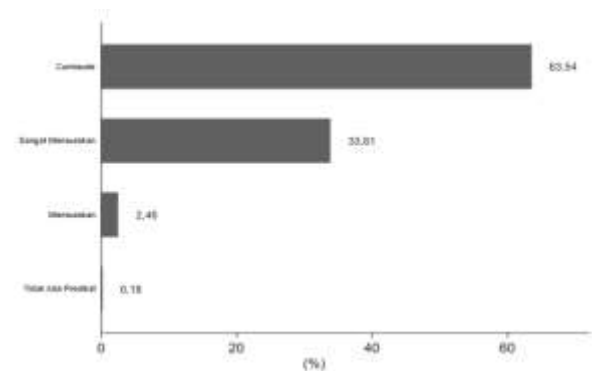Characteristics graduation predicate of UM Buton students in 2020 and 2021 are presented in Fig. 1.

Fig. 1. *Percentage Graduates Predicate of UM Buton Students*

Majority graduates get cum laude predicate as 63,54% and the very satisfying predicate as 33,81%. While the graduates to get the satisfactory predicate and didn't get the predicate were quite small, respectively 2,48% and 0,18%. Characteristics the graduation predicate according to type gender could see in Fig. 2. Majority graduate of women (67,20%) obtained predicate cum laude compared graduate of men. Likewise for the satisfactory predicate, the percentage of women graduates is greater than of men graduates.
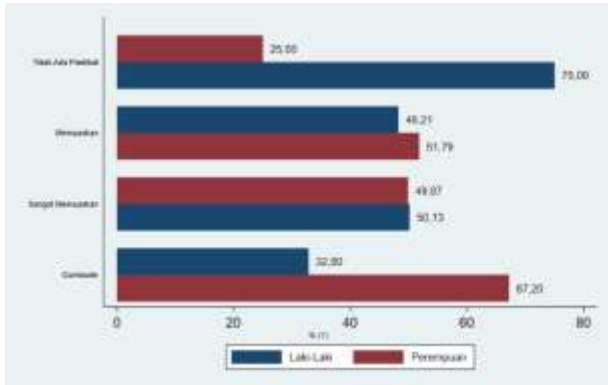
Fig. 2. *Percentage Graduates Predicate of UM Buton Students by Gender*

For the very satisfactory category, the percentage of men graduates is greater than women, although the difference is quite small. While, the graduates who didn't get the predicate were mostly men graduates. Seen from origin area (Fig. 3), percentage graduates get cum laude and very satisfying predicate from outside Baubau city the greater than graduate from inside Baubau city. While, percentage graduates get satisfying predicate from inside Baubau city the greater than from outside city.
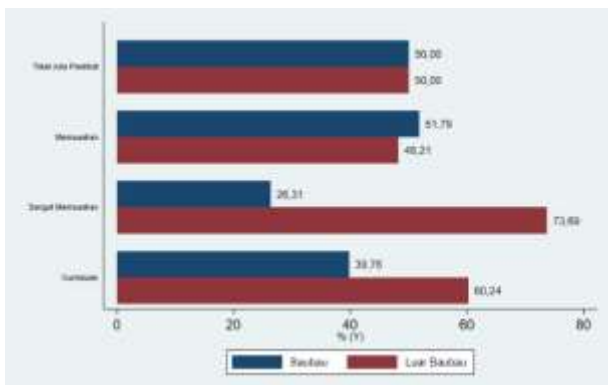

Fig. 3. *Percentage Graduates Predicate of UM Buton Students by Origin Area*
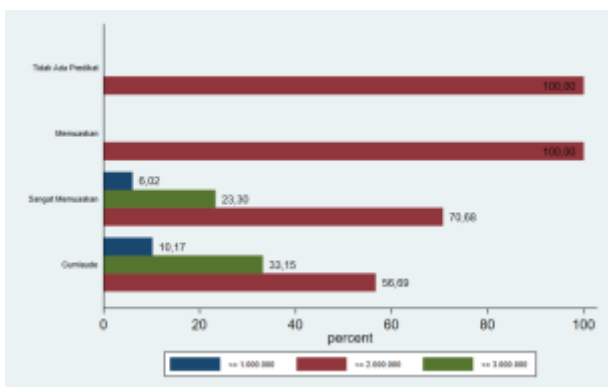

Fig. 4. *Percentage Graduates Predicate of UM Buton Students by Tuition Payment*

Based on the category of tuition payment (SPP) presented in Fig. 4, the majority of graduates get highs predicate come from study programs with tuition payment in the category of Rp. 1.000.001-Rp. 2.000.000. Likewise, all graduates don't get the predicate and get satisfactory predicate come from the study program with this category. While, at least the graduates get the predicate graduation come from study programs with tuition payment in the category $\leq$ Rp. 1.000.000.
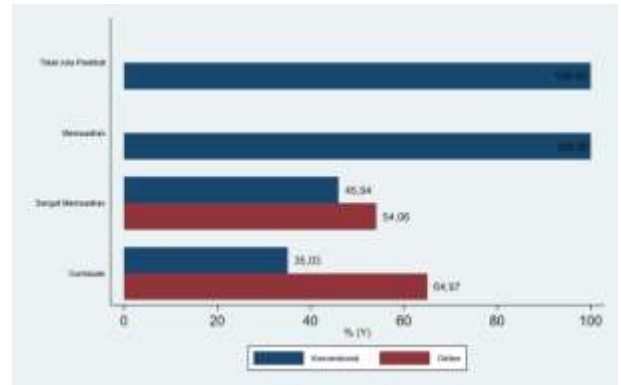

Fig. 5. *Percentage Graduates Predicate of UM Buton Students by Learning Method*

Fig. 5 presents the percentage of graduation predicate UM Buton students by the learning method used in the study program. All graduates don't get the predicate and get satisfactory predicate still use conventional method learning. In contrast to graduates get cum laude and very satisfying predicate, the biggest percentage has used *online* learning method which were respectively 64,97% and 54,06%.

3.2 Ordinal Logistics Regression Analysis

Modeling of ordinal logistics regression carried with all explanatory variable used that is gender (X_1), origin area (X_2), number graduate of the study program (X_3), category of tuition payment (X_4) and category of learning method (X_5). Next conducted to testing parameter of β by simultaneously using the LRTs test and compared with models without variable explanatory. The results LRTs test presented in Table 1.

Table 1: Test results of the **simultaneous** test

Model	LogLik	LRTs	*p-value*	Decision
Intercept Only	-1712.25			
Final	-1633.578	157.35	0.0000	H_0 rejected

Based on the results of simultaneous parameter testing, the *p-value* of 0,0000 was obtained which is smaller from score level significant $\alpha = 0,05$. So can be concluded, there is at least one parameter of β that significantly influences the graduation predicate of UM Buton students in 2020 and 2021.

Furthermore, the parameters of each explanatory variable were partially tested using the Wald test. The results of the partial test parameters are presented in Table 2, where the *p-value* obtained is 0.000 for all explanatory variables which are smaller from the significant level of $\alpha = 0,05$. So can the concluded, with using Wald's test, all parameters of explanatory variable used significant influence graduation predicate UM Buton students 2020 and 2021.

Table 2: The results of the partial test

		Coefficient	Wald	p-value	Decision
	[1]	-7.354	-13.560	0.000	H_0 rejected
	[2]	-4.598	-18.650	0.000	H_0 rejected
	[3]	-1.303	-6.300	0.000	H_0 rejected
X_1	Women	0.491	5.140	0.000	H_0 rejected
X_2	Outside Baubau	-0.601	-6.010	0.000	H_0 rejected
X_3	-	-0.007	-9.350	0.000	H_0 rejected
X_4	≤2,000,000	-0.748	-4.040	0.000	H_0 rejected
	≤3,000,000	1.792	5.850	0.000	H_0 rejected
X_5	Online	1.110	10.160	0.000	H_0 rejected

3.3 Goodness of Fit Model

The goodness of fit test is carried out to find out whether the final model that has been formed is good or not used to make predictions. With the deviance test, the model is considered good if the *p-value* is smaller than the significant level $\alpha = 0,05$. The results of the model goodness test are presented in Table 3.

Table 3: The results of goodness of fit test

	Statistics	df	p-value	Decision
Deviance	4850.453	1298	0.0000	H_0 rejected

Based on the results of testing the goodness of the model, the *p-value* of 0,0000 is obtained which is smaller than α. So can be said the ordinal logistic regression model doesn't fit the data for purposes predicting because proportion graduation predicate students who don't balanced.

3.4 Interpretation Parameter

The interpretation parameters of the ordinal logistic regression model on graduation predicate of UM Buton students using the odds ratio is presented in Table 4. Graduates UM Buton students women have opportunity to get predicate graduation more good i.e 1,355-1,971 times compared with graduate of man because values parameter $OR > 1$.

Table 4: The results of estimation odds ratio parameter

		Parameter	CL 95%	
			Lower	*Upper*
X1	Women	1.634	1.355	1971
X2	Outside Baubau	0.548	0.451	0.667
X3	-	0.993	0.992	0.994
X4	≤2,000,000	0.473	0.329	0.681
	≤3,000,000	6.03	3.292	10.947
X5	Online	3.033	2.449	3.757

While, graduates of UM Buton students who come from outside the Baubau City area are less likely to get a higher predicate because the score $OR < 1$ with confidence interval of 0,451-0,667. For graduates who come from study programs in the category of tuition payment of Rp. 2.000.001-Rp. 3.000.000 have a greater chance of get a higher predicate than other categories with score $OR = 6.003 > 1$. Likewise, graduate students who receive online-based learning have a greater chance of get higher predicate than graduates who still use conventional-based learning.

4. REFERENCES

[1] Pemerintah Republik Indonesia, *Peraturan Menteri Pendidikan dan Kebudayaan Nomor 3 Tentang Standar Nasional Pendidikan Tinggi*. 2020.

[2] M. Nusrang, R. Bakri, and A. S. Ahmar, "Analisis Regresi Logistik Ordinal Terhadap Faktor-Faktor yang Mempengaruhi Predikat Kelulusan Mahasiswa S1 Universitas Negeri Makassar," in *Proceedings of National Seminar: Research and Community Service Institute*, 2017, pp. 655–661.

[3] N. Rahmah, Indahwati, and I. D. Sulvianti, "Modelling of the Highest Level of Children's Education in the Family Using Multilevel Ordinal Logistic Regression Analysis," in *IOP Conference Series: Earth and Environmental Science*, 2018, vol. 187, no. 1, pp. 1–10.

[4] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, 2nd ed. United States: Wiley Interscience Publication, 2000.

[5] A. Agresti, *An Introduction to Categorical Data Analysis*. United Kingdom: John Wiley and Sons, 2019.