https://journal.uinsgd.ac.id/index.php/kubik/index
Vol. 10, No. 2, 2025, pp. 24-31
e-ISSN: 2686-0341 p-ISSN: 2338-0896

Logistic Regression Analysis to Determine Factors Influencing Career Choices of Undergraduate Alumni

Didi Febrian[1, *] Putri Harliana [2, b)] Nurul Ain Farhana [3, c)]

[1] *Mathematics Study Program, Universitas Negeri Medan, Indonesia*
[2] *Computer Science Study Program, Universitas Negeri Medan, Indonesia*
[3] *Statistics Study Program, Universitas Negeri Medan, Indonesia*

[*] *email: febrian.didi@unimed.ac.id*

Abstract

Undergraduate alumni generally face three career options: employment, postgraduate study, or entrepreneurship. The success of alumni careers is influenced not only by academic knowledge but also by other factors that reflect individual personal and professional abilities. These factors are often shaped through the experiences gained during undergraduate studies. The respondents of this study consisted of 100 undergraduate alumni of Universitas Negeri Medan from the graduating cohorts of 2021 and 2022. The purpose of this study was to examine the effects of GPA (X_1), study period (X_2), number of achievements (X_3), organizational activity (X_4), work intensity during undergraduate studies (X_5), and TOEFL score (X_6) on alumni career choices (Y). In logistic regression analysis, the dependent variable must have at least two categories; therefore, alumni career choices (Y) are working (0) and postgraduate study (1). The findings indicate that only two variables, work intensity during undergraduate studies (X_5) and TOEFL score (X_6) had a significant influence on alumni career choices. The logistic regression model derived from these significant predictors is expressed as $Y = -18.860 - 0.481X_5 + 0.016X_6$. The model achieved a prediction accuracy rate of 72.0%, suggesting a reasonably strong ability to classify alumni career outcomes based on the two influential factors.

Keywords: alumni career, Logistic Regression Analysis, prediction
MSC2020: 62J12

AbstraK

Alumni program sarjana pada umumnya menghadapi tiga pilihan karier, yaitu bekerja, melanjutkan studi ke jenjang pascasarjana, atau berwirausaha. Keberhasilan karier alumni tidak hanya dipengaruhi oleh pengetahuan akademik, tetapi juga oleh faktor-faktor lain yang mencerminkan kemampuan personal dan profesional individu. Faktor-faktor tersebut sering kali dibentuk melalui pengalaman yang diperoleh selama masa studi sarjana. Responden dalam penelitian ini terdiri atas 100 alumni program sarjana Universitas Negeri Medan dari angkatan kelulusan tahun 2021 dan 2022. Tujuan penelitian ini adalah untuk mengkaji pengaruh Indeks Prestasi Kumulatif (IPK) (X_1), masa studi (X_2), jumlah prestasi (X_3), keaktifan berorganisasi (X_4), intensitas bekerja selama masa studi sarjana (X_5), dan skor TOEFL (X_6) terhadap pilihan karier alumni (Y). Dalam analisis regresi logistik, variabel dependen harus memiliki minimal dua kategori; oleh karena itu, pilihan karier alumni (Y)

*) Corresponding Author
Received: 20-04-2025, Accepted: 12-12-2025, Published: 13-12-2025

diklasifikasikan menjadi bekerja (0) dan melanjutkan studi pascasarjana (1). Hasil penelitian menunjukkan bahwa hanya dua variabel, yaitu intensitas bekerja selama masa studi sarjana (X_5) dan skor TOEFL (X_6), yang berpengaruh signifikan terhadap pilihan karier alumni. Model regresi logistik yang dihasilkan dari prediktor signifikan tersebut dinyatakan sebagai $Y = -18{,}860 - 0{,}481X_5 + 0{,}016X_6$. Model ini memiliki tingkat akurasi prediksi sebesar 72,0%, yang menunjukkan kemampuan klasifikasi yang cukup kuat dalam memprediksi pilihan karier alumni berdasarkan dua faktor yang berpengaruh tersebut.

Kata kunci: karir alumni, analisis regresi logistik, prediksi
MSC2020: 62J12

Citation: D. Febrian, P. Harliana, and N. A. Farhana, "Logistic Regression Analysis to Determine Factors Influencing Career Choices of Undergraduate Alumni", KUBIK J. Publ. Ilm. Mat., Vol. 10, No. 2, pp. 24-31, 2025.

Introduction

One of the eight key performance indicators of higher education is the career success of alumni. Alumni career success serves as an important measure for universities to evaluate the quality of their graduates. In general, undergraduate alumni have three career pathways: working, postgraduate study, or entrepreneurship [1]. The success of an alumni's career is influenced not only by academic knowledge but also by other factors that reflect personal and professional abilities. These factors are often shaped by the experiences students undergo during their undergraduate studies.

Understanding the determinants of alumni career choice and success is a major concern for universities. For those pursuing employment, factors such as outstanding achievements (number of accomplishments and organizational involvement) and personality traits have been found to significantly affect career success. In contrast, academic ability, foreign language skills, and scientific writing skills show no significant effect [2]. Gender, age, study duration, wages, organizational involvement, and GPA are among the variables that influence the waiting time for graduates to get a job [3]. For alumni who wish to continue to master's programs, financial capacity becomes a crucial requirement. Scholarships provide an alternative for those with limited financial resources. To be eligible, applicants must meet three main criteria: English language proficiency, GPA requirements, and administrative completeness [4]. Among these, TOEFL scores are a universal requirement for prospective master's and doctoral students at most universities worldwide [5] and are considered an important factor in scholarship selection [6]. Entrepreneurship, however, is a less favored career path due to its inherent uncertainties, risks, and complexities in starting and sustaining a business [7]. For many in the younger generation, particularly within academic circles, entrepreneurship is not regarded as a distinctive or highly attractive career option [8].

In the field of statistics, logistic regression analysis is a widely applied method for predicting the probability of a dependent variable based on one or more independent variables. Logistic regression is particularly suitable when the dependent variable consists of two or more categories. Previous studies have applied logistic regression to identify the factors influencing the Cumulative Achievement Index (GPA) of FMIPA students at Sam Ratulangi University, Manado [9], as well as to examine the role of internal and external factors in determining academic performance [10]. In this study, logistic regression is employed to analyze the factors influencing the career choices of undergraduate alumni from Universitas Negeri Medan.

Methods

This study uses one independent variable and six dependent variables. The independent variable is Career choice (Y), while the dependent variables are GPA (X_1), study period (X_2), number of achievements (X_3), organizational activity (X_4), work intensity during undergraduate studies (X_5), and TOEFL score (X_6). Logistic regression analysis, the dependent variable must have at least two categories; therefore, alumni career choices (Y) are classified into working (0) and postgraduate studies (1). The GPA variable is measured on an interval scale with floating-point data, the length of study variable uses an interval scale with integer data, the number of achievements variable is measured on a ratio scale with integer data, organizational activities and work intensity during undergraduate studies are measured using a Likert scale, and the TOEFL score variable is measured on an interval scale with integer data.

The data used are primary data obtained through a questionnaire distributed online to undergraduate alumni of Universitas Negeri Medan in 2021 and 2022. The questionnaire contains questions about current careers and experiences gained during college such as GPA, length of study, number of achievements, organizational activity, work intensity during undergraduate, and TOEFL score. The collected data were then analyzed with the help of SPSS.

To conduct a logistic regression analysis, the researcher carried out the following stages of activities:

1. Conducting a logistic regression model suitability test using the Hosmer and Lemeshow test. The study used a significance level of 5% (0.05).
2. Conducting simultaneous parameter tests using Omnibus Tests of Model Coefficients
3. Conducting partial hypothesis tests using the Wald test statistic to test the influence of each independent variable individually in the model obtained.
4. Model interpretation based on variables that have a significant influence on the model. These variables can be identified during the Wald test.
5. Drawing conclusions with the Odds Ratio value

[11]

Results and Discussion

Data Processing

The sample used in this study was 100 undergraduate alumni of 2021 and 2022. To obtain research data, researchers distributed questionnaires online to alumni. After the data was obtained, the results of processing the alumni profile data can be seen in Table 1.

Table 1. Alumni Profile Data

Experience	Min	Max	Average
GPA	3.08	3.87	3.48
Study Period (months)	40	72	50.35
Many achievements	0	10	2.32
Organizational activity	1	5	3.25
Work Intensity during undergraduate studies	1	5	2.56
TOEFL Score	337	560	456.94

In this study, entrepreneurial careers were combined with working due to the relatively small number of entrepreneurs. Therefore, two career categories were used: working and postgraduate study. Table 2 presents the distribution of alumni career choices.

Table 2. Alumni Career Data

Alumni Career	Percentage (%)
Postgraduate Study	34
Working	66

Multicollinearity Test

Before constructing a logistic regression model, researchers conduct multicollinearity testing. Multicollinearity testing helps researchers select truly relevant predictor variables, reduces information redundancy, and improves the accuracy and reliability of the logistic regression model [12]. Table 3 shows the results of the multicollinearity test.

Table 3. Multicollinearity test

Coefficients[a]	Tolerance	VIF
Collinearity Statistics	1.000	1.000

Based on the Table 2, all independent variables (GPA, length of study, number of achievements, organizational activity, work intensity during college, and TOEFL score) have a Tolerance value of 1.000 and a VIF value of 1.000. This indicates that there is no multicollinearity among these variables, so all variables are suitable to be used in the logistic regression analysis without causing problems in the stability of the model estimation.

Logistic Regression Analysis

The logistic regression model indirectly models the independent variables based on the probabilities associated with the values on the dependent variable. To obtain a logistic regression model, several stages are carried out.

1. Conducting a logistic regression model suitability test using the Hosmer and Lemeshow test.

To test the suitability of the model, the following hypotheses are first established

H_0 : The hypothesized model fits the data.

H_1 : The hypothesized model does not fit the data.

By using Hosmer and Lemoshow test statistics, the p-value is obtained. If the sig-value <0.05 then H_0 is rejected and vice versa

Table 4. Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	3.455	8	.903

Based on Table 4, it can be seen that the p-value is marked with a Sig of 0.903. Because the sig value is greater than the research alpha (0.903>0.05), H_0 is accepted. This means that the model is suitable for the observation data, so the logistic regression model is suitable for use in the next stage.

2. Conducting simultaneous parameter tests using Omnibus Tests of Model Coefficients.

The simultaneous parameter test aims to see how the independent variable plays a role in the dependent variable simultaneously. The hypothesis used in the simultaneous parameter test is as follows

H_0 : All independent variables have no effect on the dependent variable.

H_1 : All independent variables have an effect on the dependent variable.

Simultaneous parameter testing using Omnibus Tests of Model Coefficients with the provision that if sig < 0.05 then H_0 is rejected.

Table 5. Simultaneous Parameter Tests Using

Iteration History[a,b,c]

Iteration		-2 Log likelihood	Coefficients Constant
Step 0	1	128.219	-.640
	2	128.207	-.663
	3	128.207	-.663

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	110.607	.161	.223

Omnibus Tests of Model CCoefficients

		Chi-square	df	Sig.
Step 1	Step	17.600	6	.007
	Block	17.600	6	.007
	Model	17.600	6	.007

From Table 5 (Iteration History and Model Summary), it can be seen that the -2 Log Likelihood value decreases from 128.207 at step 0 to 110.607 at step 1. This decrease indicates that the logistic regression model becomes better after including the predictors. In the Omnibus Tests of Model Coefficients, the significance value is 0.007. Since the sig < 0.05, H_o is rejected, which means that the independent variables jointly have a significant effect on the dependent variable. Furthermore, the Nagelkerke R Square value is 0.223, indicating that approximately 22.3% of the variation in career choices can be explained by the six independent variables in the model. This also implies that 77.7% of the variation in the dependent variable is influenced by other factors not included in the model, such as personal interests, family conditions, and others. In social research, a Nagelkerke R Square value between 0.2 and 0.4 is generally considered acceptable, as human behavior is highly complex and influenced by many unmeasured factors, making it rarely predictable with complete accuracy [13], [14].

3. Conducting partial hypothesis tests using the Wald test statistic.

This test is conducted to see the influence of each independent variable on the dependent variable in the model obtained. The hypothesis used in the partial hypothesis test is as follows

H_0 : The independent variable (X_i) does not affect the dependent variable

H_1 : The independent variable (X_i) affects the dependent variable

In this test, the sig value is seen in the Variables in the Equation table. If the sig value <0.05, then H_0 is rejected, the independent variable has an effect on the dependent variable.

Table 6. The Variables in The Equation Table

		β	S.E	Wald	df	Sig.	Exp(β)
Step 1	GPA (X_1)	2.867	1.645	3.039	1	.081	17.582
	Length of study (X_2)	.025	.046	.303	1	.582	1.026
	Number of achievements (X_3)	.012	.093	.018	1	.894	1.013
	Organizational activity (X_4)	.160	.209	.580	1	.446	1.173
	Work intensity during undergraduate (X_5)	-.481	.209	5.363	1	**.021**	.618
	TOEFL score (X_6)	.016	.006	6.782	1	**.009**	1.016
	Constant	-18.860	7.151	6.956	1	.008	.000

An independent variable is said to have a significant partial influence on the dependent variable if sig < 0.05. Based on Table 5, the following analysis can be carried out

- The GPA variable (X_1) has a sig value of 0.081, so sig>0.05. As a result, the GPA does not have a significant effect on alumni careers (Y).
- The length of study variable (X_2) has a sig value of 0.582, so sig>0.05. As a result, the length of study does not have a significant effect on alumni careers (Y).
- The number of achievements variable (X_3) has a sig value of 0.894, so sig>0.05. As a result, the number of achievements does not have a significant effect on alumni careers (Y).
- The organizational activity variable (X_4) has a sig value of 0.446, so sig>0.05. As a result, the organizational activity does not have a significant effect on alumni careers (Y).
- The work intensity variable during undergraduate studies (X_5) has a sig value of 0.021, so sig<0.05. As a result, the work intensity during undergraduate studies has a significant effect on alumni careers (Y).
- The TOEFL score variable (X_6) has a sig value of 0.009, so sig <0.05. As a result, the TOEFL score has a significant effect on alumni careers (Y).

Based on the results of the analysis, the factors of work intensity during undergraduate studies and TOEFL scores have a significant influence on students' career choices after graduation. This finding is consistent with the studies of Beerkens et al. and Staff and Mortimer, which emphasized that work experience during college contributes positively to the transition to the labor market and enhances employability [15], [16]. In addition, English proficiency certificates, including TOEFL, are regarded as an added value in recruitment processes, promotions, and training opportunities within international companies [17]. From an academic perspective, TOEFL scores are positively correlated with students' academic performance, thus influencing graduates' decisions to continue their education to a higher level [18].

4. Model interpretation.

The logistic regression model only uses independent variables that have a significant partial influence on the dependent variable. By considering the partial test results, the logistic regression model for alumni career choices can be written as in equation 1

$$Y = -18.860 - 0.481X_5 + 0.016X_6 \quad (1)$$

5. Drawing conclusions with the Odds Ratio value

The Odds ratio or Exp(β) value is used to facilitate interpretation of the logistic regression model. From Table 5 it can be obtained

- The odds ratio or Exp(β) value of the independent variable of work intensity during undergraduate studies is 0.618. This shows that alumni with increased work intensity during undergraduate studies are estimated to have a career choice of continuing to postgraduate studies 0.618 times greater.
- The odds ratio or Exp(β) value of the independent variable TOEFL score is 1.016. This shows that if there is an increase in the TOEFL score, it is estimated that the chance of choosing a career option of postgraduate study is 1.016 times greater.

Furthermore, to measure the accuracy of the classification, it can be seen in Table 6.

Table 7. Classification Table

			Predicted		
			Alumni Career		Percentage
Observed			Work	Postgraduate Study	Correct
Step 1	Alumni Career	Work	56	10	84.8
		Postgraduate Study	18	16	47.1
	Overall Percentage				72.0

Based on Table 7, from a sample of 66 alumni who are working, it turns out that there are only 56 alumni who are working and 10 others are studying postgraduate studies. Meanwhile, from a sample of 34 alumni who are studying postgraduate studies, it turns out that there are only 16 alumni who are studying postgraduate studies and 18 others are working. Overall, the Accuracy Percentage reaches 72.0%.

Conclusion

One measure of a university's success is reflected in the career achievements of its alumni. The logistic regression model developed in this study is expressed as: $Y = -18.860 - 0.481X_5 + 0.016\,X_6$ where X_5 represents work intensity during undergraduate studies and X_6 represents TOEFL score. This model demonstrates an overall prediction accuracy of 72%, highlighting the significant roles of work experience and English proficiency in determining alumni career paths. The findings of this study confirm that work intensity during undergraduate studies and TOEFL scores have a significant influence on students' career choices after graduation. The selection of independent variables (GPA, length of study, number of achievements, organizational involvement, work intensity during undergraduate studies, and TOEFL scores) was based on previous literature emphasizing the importance of academic, non-academic, and experiential factors in shaping career trajectories. These variables are considered relevant as they represent aspects of academic achievement (GPA, length of study, TOEFL), personal accomplishments and organizational experience, as well as work experience during college. However, it should be noted that several other variables were not included in the model, such as intrinsic motivation, family support, economic conditions, social networks, and labor market factors. These aspects may also influence alumni career choices but were beyond the scope of this study. Therefore, while the results provide valuable insights into the determinants of career choice, future research is recommended to incorporate psychosocial and contextual variables in order to develop a more comprehensive logistic regression model.

References

[1] D. Febrian, P. Harliana, and M. As, "Career Prediction of Undergraduate Alumni Using Fuzzy Logic," in *The 10th Annual International Seminar on Trends in Science and Science Education (AISTSSE) 2023*, Sciendo, 2024, pp. 458–466. doi: 10.2478/9788367405782-057.

[2] Z. S. Syafhil and L. Kartika, "Analisis Kesuksesan Karir Alumni Program Mahasiswa Berprestasi di Perguruan Tinggi," *Perspektif Ilmu Pendidikan*, vol. 35, no. 1, pp. 9–24, Apr. 2021, doi: 10.21009/pip.351.2.

[3] Y. Yulmiati, I. H. Rahmi, and F. Yanuar, "Klasifikasi Masa Tunggu Kerja Lulusan S1 Matematika Unand Menggunakan Naive Bayes Classifier," *J. Ris. & Ap. Mat*, vol. 5, no. 1, pp. 29–37, 2021.

[4] S. Ariawan *et al.*, "Mentoring Beasiswa Dan Studi Lanjut Secara Online Melalui Program Sekolah Beasiswa Straya Untuk Pelamar Beasiswa Pascasarjana Tujuan Dalam Dan Luar Negeri," *JMM (Jurnal Masyarakat Mandiri)*, vol. 5, no. 5, pp. 2165–2175, 2021, doi: 10.31764/jmm.v5i5.5210.

[5] R. Herwandar, D. A. Safryono, and P. Y. Haryono, "Evaluasi Program Matrikulasi 'TOEFL' Mahasiswa Universitas Al Azhar Indonesia 2010/2011," *Jurnal AL-AZHAR INDONESIA SERI HUMANIORA*, vol. 1, no. 3, pp. 179–188, 2012.

[6] S. Amalia, H. Husnaini, and R. Ilma, "Pelatihan Pengenalan TOEFL Sebagai Persiapan Studi Lanjut Pada Siswa-Siswi MAN 2 Palembang," *SELAPARANG: Jurnal Pengabdian Masyarakat Berkemajuan*, vol. 7, no. 3, pp. 1599–1604, 2023.

[7] G. Segal, D. Borgia, and J. Schoenfeld, "The motivation to become an entrepreneur," *International Journal of Entrepreneurial Behaviour and Research*, vol. 11, no. 1, pp. 42–57, 2005, doi: 10.1108/13552550510580834.

[8] S. S. Thrikawala, "The Determinants of Entrepreneurial Intention among Academics in Srl Lanka," in *International Conference on Economics and Finance Research IPEDR vol.4*, Singapore: IACSIT Press, 2011, pp. 454–458.

[9] A. Y. Tampil, H. Komalig, and Y. Langi, "Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado," *d\'Cartesian: Jurnal Matematika Dan Aplikasi,* vol. 6, no. 2, pp. 56–62, 2017.

[10] M. Marna, M. Saftari, P. Jana, and M. Maxrizal, "Analisis Regresi Logistik Biner Untuk Memprediksi Faktor Internal Dan Eksternal Terhadap Indeks Prestasi," *Delta: Jurnal Ilmiah Pendidikan Matematika*, vol. 9, no. 1, p. 47, Jan. 2021, doi: 10.31941/delta.v9i1.1251.

[11] A. Mariani, A. Sauddin, A. Adiatma, and H. Nufus, "Analisis Regresi Logistik Biner Untuk Mengidentifikasi Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif Mahasiswa Matematika Universitas Islam Negeri Alauddin Makassar," *Jurnal Matematika dan Statistika serta Aplikasinya*, vol. 11, no. 1, pp. 6–11, 2023.

[12] S. L. R. X. S. David W. Hosmer Jr., *Applied Logistic Regression*. John Wiley & Sons, Inc., 2013.

[13] M. Shah, "Could someone explain me about Nagelkerke R Square in Logit Regression analysis? ," Sep. 2023. Accessed: Apr. 15, 2025. [Online]. Available: https://www.researchgate.net/post/Could_someone_explain_me_about_Nagelkerke_R_Square_in_Logit_Regression_analysis

[14] Sarah Lee, "A Comprehensive Guide to McFadden's R-squared in Logistic Regression," https://www.numberanalytics.com/blog/comprehensive-guide-mcfaddens-r-squared-logistic-regression. Accessed: Apr. 15, 2025. [Online]. Available: https://www.numberanalytics.com/blog/comprehensive-guide-mcfaddens-r-squared-logistic-regression

[15] M. Beerkens, E. Mägi, and L. Lill, "University studies as a side job: Causes and consequences of massive student employment in Estonia," *High Educ (Dordr)*, vol. 61, no. 6, pp. 679–692, Jun. 2011, doi: 10.1007/s10734-010-9356-0.

[16] J. Staff and J. T. Mortimer, "Educational and work strategies from adolescence to early adulthood: consequences for educational attainment," *Social Forces*, vol. 85, no. 3, pp. 1169–1194, Mar. 2007, doi: 10.1353/sof.2007.0057.

[17] Y. C. Pan and C. Roever, "Consequences of test use: a case study of employers' voice on the social impact of English certification exit requirements in Taiwan," *Language Testing in Asia*, vol. 6, no. 1, Dec. 2016, doi: 10.1186/s40468-016-0029-5.

[18] A. Pranita Devi, "The Relationship between English Proficiency and Academic Achievement of Indonesian EFL Postgraduate Students," *Journal of English Language Learning (JELL)*, vol. 7, no. 1, pp. 303–308, [Online]. Available: https://www.researchgate.net/publication/390244510