PARAMETER
JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA

e-ISSN. 2829-6303
Vol. 05, No 2, September 2026
Pages 249-266

A Statistical Insight into Graduate Income: An IRLS-Optimized Generalized Linear Modelling Approach

Yusma Yanti [1,2], **Hutomo Atman Maulana** [1,3*], **Salmun K Nasib** [1,4], **Anang Kurnia** [1]

[1]Statistics and Data Science Study Program, IPB University
Raya Dramaga St., Kabupaten Bogor, 16680, Jawa Barat, Indonesia
[2]Computer Science Study Program, Pakuan University
Pakuan St., Kota Bogor, 16143, Jawa barat, Indonesia
[3]International Business Administration Study Program, Politeknik Negeri Bengkalis
Bathin Alam St., Kabupaten Bengkalis, Riau, 28711, Riau, Indonesia
[4]Statistics Study Program, Gorontalo State University
Prof. Dr. Ing. B. J. Habibie St., Kabupaten Bone Bolango, Gorontalo, 96119, Indonesia

E-mail Correspondence Author: *hutomomaulana@gmail.com

Abstract

This study examines the statistical determinants of graduate income using Maximum Likelihood Estimation implemented through the Iteratively Reweighted Least Squares algorithm. A generalized linear model with a logit link was estimated to assess how technical competence, English proficiency, and tuition-financing type influence the probability of earning above the regional minimum wage. The IRLS procedure enabled iterative refinement of coefficient estimates and standard errors, with convergence achieved as successive updates stabilized, indicating a well-behaved likelihood surface. Results show that academic competence and English proficiency significantly increase the log-odds of higher income, highlighting the importance of human-capital attributes in early-career outcomes. Study duration also exhibits a positive and statistically significant association with the likelihood of earning above the minimum wage. In contrast, tuition-financing type shows a positive but statistically non-significant effect, suggesting that its influence on graduate income requires further investigation. Although the model identifies several meaningful determinants, its moderate discriminative capacity indicates that pre-graduation characteristics alone cannot fully explain income variation. By explicitly incorporating MLE with IRLS, this study provides a transparent statistical framework for analyzing graduate employability outcomes and contributes empirical evidence on factors associated with early-career income attainment. The findings encourage future research integrating labor-market, employer, and longitudinal career information to improve predictive performance.

Keywords: *Graduate employability, binary outcome modelling, maximum likelihood estimation.*

: https://doi.org/10.30598/parameterv5i2pp249-266

Submitted: December 2025
Accepted: September 2026

1. INTRODUCTION

The growing global emphasis on graduate employability has intensified the need for rigorous analytical frameworks capable of identifying the determinants of labor-market outcomes in higher education systems. Universities today are increasingly evaluated not only for academic excellence but also for their effectiveness in preparing graduates to transition into competitive professional environments. As labor markets evolve toward digitalization, automation, and global mobility, employers place increasing value on adaptivity, communication skills, and digital competence, reshaping what constitutes employability. These transformations highlight the importance of quantitative, evidence-based research capable of explaining how diverse pre-graduation characteristics influence early-career outcomes. Recent studies underscore that employability is multidimensional, shaped by academic performance, competencies, institutional context, and demographic factors, which collectively contribute to employment probabilities and income trajectories [1], [2], [3]. The rapid expansion of tracer studies in Indonesia and internationally further underscores the need for advanced modeling techniques capable of extracting meaningful insights from increasingly rich graduate datasets.

Tracer studies have become a central component of institutional monitoring and quality assurance, offering systematic and longitudinal insights into graduates' employment status, income levels, job relevance, and perceived competence. The dataset employed in this study derives from the 2023 tracer study of Universitas Pakuan, which provides a comprehensive profile of alumni experiences within one year of graduation. Similar datasets have been instrumental in informing policy, benchmarking performance, and evaluating curriculum alignment with industry needs. However, the statistical structure of tracer study data frequently comprising binary, categorical, or ordinal response variables poses methodological challenges when analyzed using classical linear regression. Such models impose assumptions of normality, homoscedasticity, and linearity that are often violated in employability research. In contrast, generalized linear models (GLMs), which extend classical regression by enabling alternative link functions and distribution families, provide a flexible and theoretically robust approach for analyzing heterogeneous response types, as increasingly demonstrated in higher-education analytics literature [4], [5].

Parameter estimation is a critical component of GLM implementation, particularly for binary outcomes requiring link functions such as the logit. Maximum Likelihood Estimation (MLE) is the standard technique for estimating GLM parameters; however, the corresponding likelihood equations rarely have closed-form solutions. As a result, iterative numerical optimization procedures are required. Among these, the Iteratively Reweighted Least Squares (IRLS) algorithm is widely recognized for its computational efficiency, convergence stability, and suitability for logistic models. IRLS refines parameter estimates through repeated weighted least-squares updates, enabling the model to account for the variance structure inherent in binary data. Studies in computational statistics continue to highlight the superior convergence properties of IRLS in logistic regression, particularly when dealing with medium to large sample sizes [6], [7], [8]. Thus, embedding IRLS within an MLE framework is essential for ensuring the accuracy and reliability of statistical inference in graduate employability studies.

The core research problem addressed in this study concerns the limited understanding of how combinations of academic, demographic, and competency variables influence the probability of earning income above the regional minimum wage. While prior research has often focused on employment status or job relevance, fewer studies have systematically investigated income differentiation as a function of multidimensional graduate characteristics. Moreover, much existing work in the Indonesian context remains descriptive or constrained by traditional regression assumptions. This creates a methodological gap: higher-education leaders require predictive tools that not only identify significant predictors but also provide interpretable measures such as odds ratios and estimated probabilities under valid distributional assumptions.

A review of international literature reveals that graduate income is influenced by field of study, gender, academic performance, study duration, and a variety of behavioral competencies. Research across multiple regions has found persistent disciplinary income gaps, gender-related wage disparities, and moderate effects of academic achievement on earnings [9], [10], [11]. Meanwhile, competencies such as communication, teamwork, ethics, and digital literacy are increasingly recognized as salient determinants of early career success, particularly in technologically dynamic sectors [12], [13]. The literature also indicates that financial factors, particularly scholarship status, play a meaningful role in facilitating academic engagement and improving career opportunities; scholarship recipients often demonstrate stronger labor-market outcomes due to both selection mechanisms and the reduced financial burden associated with study [14]. However, few studies integrate these multidimensional predictors into a single statistical model capable of estimating their relative contributions with methodological rigor. Within the GLM domain, logistic regression is frequently used to analyze binary employability outcomes, but many prior studies focus on limited sets of predictors or fail to articulate the estimation methods used. Without proper estimation procedures or convergence diagnostics, models may yield unstable or biased estimates, reducing their utility for policymaking. A statistically rigorous application of GLM-MLE using IRLS therefore fills a methodological gap by providing transparent, replicable, and robust parameter estimation.

Against this backdrop, this study employs a binary GLM estimated using the IRLS algorithm to examine the probability that graduates from Universitas Pakuan's 2023 cohort earn income above the regional minimum wage. The model incorporates multiple predictors i.e. faculty, gender, study funding type, study duration, academic ability, ethics, English proficiency, IT skills, communication, teamwork, and self-development those yielding a multidimensional view of income determinants. The novelty of this study lies not in the estimation procedure itself, which follows the standard GLM-MLE framework, but in its application to a comprehensive graduate tracer-study dataset from a private university in Indonesia. By combining demographic characteristics, academic factors, scholarship status, and self-reported competencies within a unified analytical framework, this study provides empirical evidence on graduate income determinants in a higher-education setting that remains underrepresented in the international employability literature. The findings contribute to a better understanding of how graduate attributes relate to early-career income outcomes in private-university contexts while providing empirically grounded insights for curriculum refinement, scholarship policy, and institutional performance evaluation.

2. METHOD

2.1. Data Sources and Research Variables

The data used in this study originate from the 2023 Tracer Study of Universitas Pakuan, an institution-wide survey designed to capture comprehensive information on alumni employment outcomes within one year after graduation. The response variable in this study is defined with reference to the official regional wage regulation issued by the Government of West Java Province. Specifically, the income threshold distinguishing the binary outcome is based on *Surat Keputusan Gubernur Nomor 561.7/Kep.766-Kesra/2022 tentang Upah Minimum Kabupaten/Kota di Daerah Provinsi Jawa Barat Tahun 2023*, which stipulates that the minimum monthly wage (UMR) for the Bogor region in 2023 is Rp 4.639.429. Accordingly, the binary response is coded as 1 if a graduate's self-reported income exceeds this regulatory threshold and 0 otherwise. Predictor variables include academic and demographic characteristics as well as self-reported competencies that align with existing evidence on graduate employability determinants [15]. **Table 1** summarizes the variables used in the analysis, including their operational definitions and measurement scales.

Table 1. Research variables

Variable	Description	Scale
Income (Response (Y))	Above UMR = 1; Below UMR = 0	Binary
Faculty	FMIPA (X_1), FISIB (X_2), FT (X_3), FKIP (X_4), FEB (baseline)	Nominal
Gender (X_5)	Male, Female (baseline)	Nominal
Scholarship (X_6)	No (0), Yes (1)	Nominal
Years of Study (X_7)	Year (4~7)	Ratio
Ethics (X_8)	Self-reported competency	Ordinal (1–5)
Academics (X_9)	Self-reported competency	Ordinal (1–5)
English (X_{10})	Self-reported competency	Ordinal (1–5)
IT (X_{11})	Self-reported competency	Ordinal (1–5)
Communication (X_{12})	Self-reported competency	Ordinal (1–5)
Teamwork (X_{13})	Self-reported competency	Ordinal (1–5)
Self-Development (X_{14})	Self-reported competency	Ordinal (1–5)

Source: Pakuan University Tracer Study (2023)

2.2. Data Analysis Procedure

The methodology used in this study consists of three core stages: data preprocessing, model estimation using Maximum Likelihood Estimation (MLE) with Iteratively Reweighted Least Squares (IRLS), and model performance evaluation.

2.2.1. Data Preprocessing

The preprocessing stage involved examining the structure of the dataset, identifying missing values, detecting duplicate records, and screening for invalid entries. Ensuring data integrity is essential for the reliability of MLE-based estimation, as likelihood optimization can be highly sensitive to outliers and inconsistent coding. Missing values were assessed for patterns and mechanism, and invalid entries were corrected or removed based on standard data-cleaning procedures. Duplicate observations often arising from multiple survey submissions were identified using alumni's name.

2.2.2. Modelling

The binary nature of the response variable motivated the use of a generalized linear model (GLM) with a Bernoulli distribution and logit link function. In this study, the random component assumes that each response $Y_i \sim \text{Bernoulli}(p_i)$, where $p_i = P(Y_i = 1)$ denotes the probability of earning above the minimum wage (UMR), the probability mass function is.

$$f(y_i \mid p_i) = p_i^{y_i}(1 - p_i)^{1-y_i} \tag{1}$$

the expectation $E(Y_i) = p_i$ and variance are $Var(Y_i) = p_i(1 - p_i)$ the systematic component is expressed as.

$$\eta_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i} + \beta_8 X_{8i} + \beta_9 X_{9i} + \beta_{10} X_{10i} + \beta_{11} X_{11i} + \beta_{12} X_{12i} + \beta_{13} X_{13i} + \beta_{14} X_{14i} \quad (2)$$

while the logit link function

$$g(p_i) = \log(\frac{p_i}{1 - p_i}) = \eta_i \quad (3)$$

the inverse link is

$$g^{-1}(\eta_i) = p_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad (4)$$

assuming independence across observations, the likelihood of $\boldsymbol{\beta}$ is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} p_i^{y_i}(1 - p_i)^{1-y_i} \quad (5)$$

substituting $p_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}$, gives

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{e^{\eta_i}}{1 + e^{\eta_i}}\right)^{y_i} \left(\frac{1}{1 + e^{\eta_i}}\right)^{1-y_i} = \prod_{i=1}^{n} \left(\frac{e^{y_i\eta_i}}{1 + e^{\eta_i}}\right) \quad (6)$$

taking the natural logarithm of the likelihood yields

$$\mathrm{l}(\boldsymbol{\beta}) = \log L\,(\boldsymbol{\beta}) = \sum_{i=1}^{n} [y_i\eta_i - \log(1 + e^{\eta_i})] \quad (7)$$

the first derivative (Score Function) is

$$\boldsymbol{U}(\boldsymbol{\beta}) = \frac{\partial \mathrm{l}(\beta)}{\partial \beta} = \sum_{i=1}^{n} x_i(y_i - \mu_i) = \boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{\mu}) \quad (8)$$

the second derivative of the log-likelihood (Observed Information) is

$$\boldsymbol{H}(\beta) = -\frac{\partial^2 \mathrm{l}(\beta)}{\partial \beta \partial \beta^\top} = \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X} \quad (9)$$

where

$$\boldsymbol{W} = \mathrm{diag}\big(p_i(1 - p_i)\big) \quad (10)$$

The Fisher Information is defined as the expectation of the observed information:

$$\boldsymbol{I}(\beta) = E\big(\boldsymbol{H}(\beta)\big) = \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X} \quad (11)$$

The IRLS updating equation for maximizing the log-likelihood is

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \big(\boldsymbol{X}^\top \boldsymbol{W}^{(t)} \boldsymbol{X}\big)^{-1} \boldsymbol{X}^\top\big(\boldsymbol{y} - \boldsymbol{\mu}^{(t)}\big) \quad (12)$$

using the working dependent variable

$$\boldsymbol{z}^{(t)} = \boldsymbol{\eta}^{(t)} + \frac{\boldsymbol{y} - \boldsymbol{\mu}^{(\boldsymbol{t})}}{\boldsymbol{p}^{(t)}(1 - \boldsymbol{p}^{(t)})} \quad (13)$$

the IRLS step becomes a weighted least-squares update

$$\boldsymbol{\beta}^{(t+1)} = \big(\boldsymbol{X}^\top \boldsymbol{W}^{(t)} \boldsymbol{X}\big)^{-1} \boldsymbol{X}^\top \boldsymbol{W}^{(t)} \boldsymbol{z}^{(t)} \quad (14)$$

The IRLS algorithm iterates until

$$|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}| < 10^{-5} \tag{15}$$

2.2.3. Model Performance Evaluation

Model adequacy was assessed through multiple diagnostic techniques. First, the Hosmer–Lemeshow goodness-of-fit test was used to evaluate the correspondence between predicted and observed probabilities across deciles of risk. A non-significant Hosmer–Lemeshow statistic indicates that the model fits the data well. Additionally, the residual deviance was evaluated against the null deviance to determine the overall improvement in fit attributable to the inclusion of predictors. Lower deviance values indicate better model performance, and the chi-square approximation enables hypothesis testing regarding model adequacy [16].

To evaluate discriminative ability, the Receiver Operating Characteristic (ROC) curve was constructed. The ROC curve plots the true-positive rate against the false-positive rate across varying probability thresholds. The Area Under the Curve (AUC) quantifies the overall classification accuracy, where values closer to 1 indicate superior discriminative performance. ROC-based evaluation has become a standard complementary tool to goodness-of-fit statistics, particularly in classification models involving policy-relevant binary outcomes [17].

3. RESULTS AND DISCUSSION

3.1. Data Description

Descriptive statistics from response dan predictor variables are shown in Table 2. The income variable, coded as a binary outcome, shows that 69.67% of graduates earn below the regional minimum wage, while 30.33% earn above it. This indicates that the majority of respondents fall into the lower-income category, thereby underscoring the importance of identifying determinants that influence the probability of higher earnings. The dataset exhibits a mild class imbalance with a minority class. Mild imbalance (20-40%) generally does not necessitate specialized resampling techniques such as SMOTE, as standard logistic regression models remain robust in this range [18]. Furthermore, the number of events per predictor variable exceeds the recommended threshold of 10, ensuring the stability of the parameter estimates [19].

In terms of faculty distribution, respondents originate primarily from FISIB (31.12%) and FMIPA (22.43%), followed by FKIP (24.64%), FEB as the baseline category (13.74%), and FT (8.06%). The gender composition reflects a predominance of female graduates (64.77%), with male respondents accounting for 35.23%. Scholarship holders represent only 5.85% of the sample, while 94.15% financed their studies independently, suggesting limited penetration of financial assistance among students. The average study duration is 5.16 years, indicating that most graduates complete their programs slightly beyond the standard four-year period.

Competency-related ordinal variables exhibit distributions skewed toward higher skill levels. Ethics displays a concentration at levels 4 and 5, representing 39.81% and 50.39% of respondents, respectively, with a mean score of 4.39. Academic skills follow a similar pattern, where 42.02% of respondents fall at level 3, 33.81% at level 4, and 14.22% at the highest level, yielding a mean of 3.52. Information technology skills show 49.92% at level 4 and 37.28% at level 5 (mean 4.24), demonstrating relatively strong digital competence within the cohort.

Table 2. Descriptive statistics of response and predictor variables

Variable	Value	Freq.	%	Mean or Mode
Income (Y)	< UMR (0)	441	69.67	0
	>= UMR (1)	192	30.33	
Faculty	FMIPA (X_1)	142	22.43	FISIB
	FISIB (X_2)	197	31.12	
	FT (X_3)	50	8.06	
	FKIP (X_4)	156	24.64	
	FEB (baseline)	87	13.74	
Gender (X_5)	Male (1)	223	35.23	0
	Female (0)	410	64.77	
Scholarship (X_6)	No (0)	596	94.15	0
	Yes (1)	37	5.85	
Years of Study (X_7)	Year (4~7)	-	-	5.16
Ethics (X_8)	1 (Very low)	6	0.95	4.39
	2	0	0	
	3	56	8.85	
	4	252	39.81	
	5 (Very high)	319	50.39	
Academics (X_9)	1 (Very low)	4	0.63	3.52
	2	59	9.32	
	3	266	42.02	
	4	214	33.81	
	5 (Very high)	90	14.22	
English (X_{10})	1 (Very low)	0	0	4.15
	2	8	1.26	
	3	91	14.38	
	4	329	51.97	
	5 (Very high)	205	32.39	
IT (X_{11})	1 (Very low)	0	0	4.24
	2	0	0	
	3	81	12.80	
	4	316	49.92	
	5 (Very high)	236	37.28	
Communication (X_{12})	1 (Very low)	4	0.63	4.25
	2	5	0.79	
	3	98	15.48	
	4	246	38.86	
	5 (Very high)	280	44.23	
Teamwork (X_{13})	1 (Very low)	1	0.16	4.39
	2	2	0.32	
	3	65	10.27	
	4	247	39.02	
	5 (Very high)	318	50.24	
Self-Development (X_{14})	3	50	7.90	4.49
	4	220	34.76	
	5 (Very high)	363	57.35	

Source: Pakuan University Tracer Study (2023)

Communication skills also cluster at higher levels, with 38.86% at level 4 and 44.23% at level 5 (mean 4.25). Teamwork skills are similarly concentrated at levels 4 (39.02%) and 5 (50.24%), yielding a mean of 4.39, which highlights strong collaborative abilities among graduates. Self-development scores reinforce this pattern, with 34.76%

of respondents at level 4 and 57.25% at level 5, resulting in a mean of 4.49. English proficiency, while slightly more dispersed, remains weighted toward upper categories, with 32.39% at level 4 and 32.39% at level 5, and a corresponding mean of 4.15.

3.2. Modelling

3.2.1. Checking the assumptions

Assessing the Pearson correlations among predictors is an essential preliminary step because it allows the analyst to identify potential linear dependencies that may compromise coefficient stability and interpretability in a generalized linear model. The correlation matrix provides a concise summary of the linear relationships among the predictors included in the logistic regression model, enabling an initial assessment of potential multicollinearity and shared variance structures across competencies. The correlation matrix summarizes the strength and direction of linear associations among the predictors, providing an initial check on whether any variables exhibit relationships strong enough to affect the stability of the generalized linear model. **Figures 1** shows that several competency measures most notably Communication, Teamwork, and IT have moderate positive correlations (e.g., 0.42–0.55), implying that these skills tend to co-develop within respondents. In contrast, variables such as Duration of Study (Year) and Academics exhibit correlations very close to zero, indicating minimal shared variability and suggesting that they contribute independent information to the model. Overall, the numerical values confirm that none of the predictor pairs approach the high correlation thresholds typically associated with multicollinearity, supporting their suitability for inclusion in the GLM.

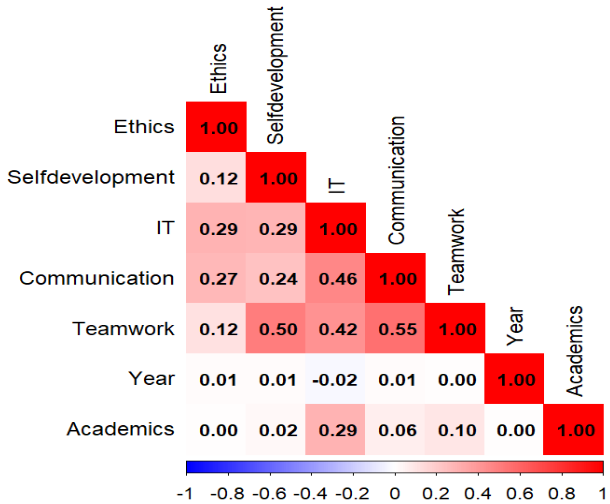

Figure 1. Pearson Correlation for Predictor Variable

A series of model diagnostics was performed to evaluate the adequacy and robustness of the fitted binary logistic regression model. The visual summaries presented in Figure 2 provide insights into two essential diagnostic dimensions: (i) the presence of influential observations, and (ii) potential multicollinearity among predictors. Both diagnostics are critical for ensuring the stability of parameter estimates and the reliability of subsequent inference. The left panel illustrates the leverage residual plot, which assesses influential observations based on standardized residuals plotted against leverage values. The majority of data points lie well within the 0.5 and 0.9 Cook's distance contours, indicating the absence of observations that exert disproportionate influence on the fitted model. Only a small number of cases appear near the outer envelope, yet still remain below conventional influence thresholds. This pattern suggests that the model is not unduly driven by outliers or extreme leverage points, thereby supporting the stability of the coefficient estimates.

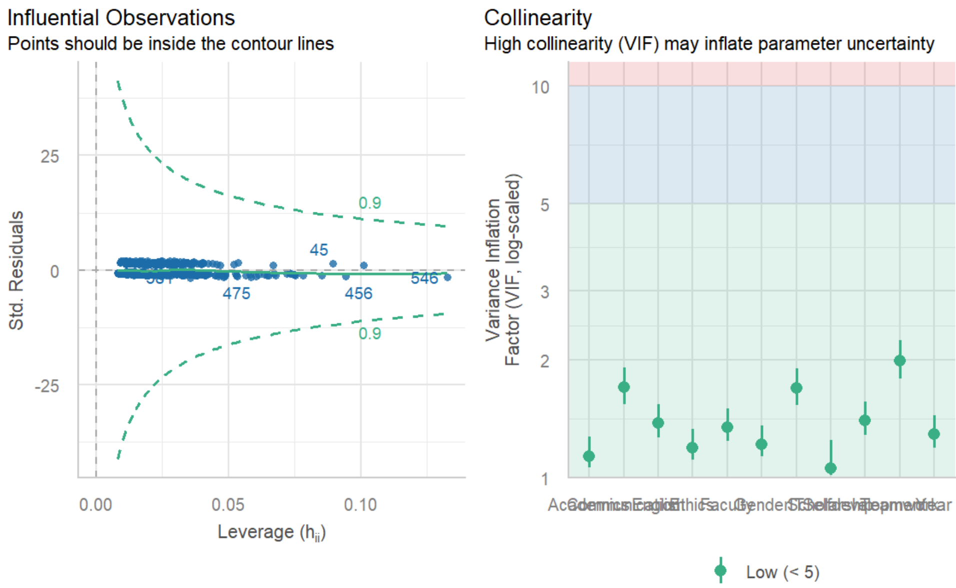

Figure 2. Diagnostic Plots for the Binary Generalized Linier Model

The right panel presents the Variance Inflation Factor (VIF) values for the predictors, displayed on a log-scaled axis to facilitate comparative assessment. All VIF values fall below the commonly accepted threshold of 5, signifying low levels of multicollinearity among the independent variables. This confirms that the predictors contribute distinct information to the model and that multicollinearity is unlikely to inflate standard errors or compromise the interpretability of coefficient estimates. Collectively, the diagnostic results indicate that the model satisfies key assumptions related to influence and predictor independence. The absence of high-leverage outliers and the low VIF values reinforce the reliability of the model and justify proceeding to substantive interpretation of parameter estimates and predicted effects.

3.2.2. Parameter estimation and odds ratio

The estimation results were obtained through the application of a binary logistic regression model using the Generalized Linear Model (GLM) framework with a logit link function. Prior to model fitting, all variables were systematically prepared by coding the income outcome as a binary response and specifying the appropriate measurement scales for nominal, ordinal, and continuous predictors. Parameter estimation was conducted using Maximum Likelihood Estimation with IRLS methods, producing coefficient estimates, standard errors, and associated significance levels for each predictor. This output can be seen in the Table 2. The estimation output indicates that only a subset of the predictors makes a significant contribution to the log-odds of earning income above the regional minimum wage. The variables representing Academics and English show statistically significant positive effects, suggesting that higher competency levels in these domains increase the likelihood of attaining higher income after graduation. Meanwhile, Years also demonstrates a positive and significant association, indicating that extended study duration may reflect accumulated academic or experiential advantages that translate into improved income outcomes.

Table 3. Coefficient Estimates for Predictors with IRLS

Variable	Iteration						
	1	2	3	4	5*		
	$(\hat{\beta})$	$(\hat{\beta})$	$(\hat{\beta})$	$(\hat{\beta})$	$(\hat{\beta})$	SE $(\hat{\beta})$	p-value
Intercept (β_0)	-3.2673	-3.9832	-4.0318	-4.0320	-4.0320	1.0955	**0.00023****
FT (β_1)	0.0785	0.0936	0.0949	0.0949	0.0949	0.4082	0.81619
FMIPA(β_2)	0.1619	0.2004	0.2028	0.2028	0.2028	0.3072	0.50910
FISIB (β_3)	0.1570	0.1971	0.1997	0.1997	0.1997	0.2897	0.49059
FKIP (β_4)	0.1293	0.1520	0.1528	0.1528	0.1528	0.3202	0.63317
Scholarship (β_5)	0.5275	0.6161	0.6219	0.6219	0.6219	0.3649	0.08828
Year (β_6)	0.1621	0.1959	0.1978	0.1978	0.1978	0.0962	**0.03974****
Male (β_7)	0.2593	0.3123	0.3155	0.3156	0.3156	0.2024	0.11895
Ethics (β_8)	-0.0314	-0.0268	-0.0259	-0.0259	-0.0259	0.1326	0.84516
Academics (β_9)	0.2396	0.2972	0.3010	0.3010	0.3010	0.1107	**0.00656****
English (β_{10})	0.3682	0.4643	0.4714	0.4714	0.4714	0.1549	**0.00234****
IT (β_{11})	0.1629	0.2052	0.2088	0.2088	0.2088	0.1789	0.24304
Communication (β_{12})	0.0067	0.0143	0.0152	0.0152	0.0152	0.1493	0.91898
Teamwork (β_{13})	-0.1849	-0.2511	-0.2570	-0.2570	-0.2570	0.1797	0.15267
Selfdevelopment (β_{14})	-0.1657	-0.1997	-0.2018	-0.2018	-0.2018	0.1621	0.12324
Log-likelihood	-438.762	-369.771	-368.539	-368.534		-368.534	

*Convergent with error 4.395478e-09

**Significant with $\alpha = 0.05$

In contrast, faculty affiliation, gender, and most competency variables do not exhibit statistical significance, implying that their direct influence on income status is limited within this sample. The coefficient for Scholarship approaches marginal significance, which may suggest a potential trend that scholarship recipients have enhanced employment outcomes, although this effect remains inconclusive. The model intercept is negative and significant, reflecting the low baseline probability of earning above the minimum wage in the absence of predictor effects. Overall, these findings highlight the selective importance of specific academic and skill-related attributes in shaping graduates' economic attainment, while underscoring the need to interpret nonsignificant predictors cautiously within the broader context of labor market heterogeneity.

$$
\begin{aligned}
ln\left(\frac{p_i(x)}{1-p_i(x)}\right) &= \eta_i \\
&= -4.032 + 0.09489X_{1i} + 0.20282X_{2i} + 0.19968X_{3i} \\
&\quad + 0.15284X_{4i} + 0.62192X_{5i} + 0.19782X_{6i} + 0.31556X_{7i} \\
&\quad - 0.02591X_{8i} + 0.301X_{9i} + 0.47139X_{10i} + 0.20882X_{11i} \\
&\quad + 0.01518X_{12i} - 0.25701X_{13i} - 0.20183X_{14i}
\end{aligned}
$$

In this model, the relationship between the predictors and the binary outcome is expressed through the odds of the event occurring, rather than through direct probabilities. Odds represent the ratio between the probability of success and the probability of failure, providing a multiplicative scale that is suitable for modeling nonlinear probability changes. The logistic model assumes that each predictor has a constant multiplicative effect on the odds, which becomes additive on the log-odds scale. Consequently, exponentiated coefficients from the model correspond to odds ratios, indicating how the odds of achieving income above the regional minimum wage increase or decrease for a one-unit change in a predictor, holding all other variables constant. The odds ratio for predictor j is shown in in **Equation (2)**.

$$OR_j = e^{\beta_j} \tag{2}$$

Based on the model so the odds ratio for each predictor can be seen in Table 3. The results from Table 3 indicate that only a subset of predictors demonstrates a significant effect on the likelihood of earning income above the regional minimum wage. Study duration shows a positive and significant association, with an odds ratio of 1.22, implying that each additional year of study increases the odds of higher income by approximately 21.9%. Academics skill competency also exhibits a meaningful influence, where a one-unit improvement increases the odds of higher income by 35.1%. English proficiency emerges as the strongest predictor, with an odds ratio of 1.60, indicating a 60.2% increase in the odds of achieving higher income for each incremental improvement in proficiency level. Although the study financing variable (scholarship vs. self-funded) does not reach statistical significance, its positive coefficient suggests a potential tendency for scholarship recipients to have better income outcomes, albeit with insufficient evidence to confirm this effect within the present sample. Overall, the findings highlight the selective importance of academic and skill-based factors in shaping graduates' income attainment.

Table 3. Odds Ratio for Each Predictor

Predictor	OR	CI Lower	CI Upper	p-value
Intercept (β_0)	0.018	0.002	0.149	0.0002*
FT (β_1)	1.100	0.490	2.442	0.8162
FMIPA(β_2)	1.225	0.674	2.255	0.5091
FISIB (β_3)	1.221	0.697	2.176	0.4906
FKIP (β_4)	1.165	0.625	2.201	0.6332
Scholarship (β_5)	1.862	0.899	3.795	0.0883
Year (β_6)	1.219	1.009	1.472	0.0397*
Male (β_7)	1.371	0.921	2.038	0.1190
Ethics (β_8)	0.974	0.751	1.269	0.8452
Academics (β_9)	1.351	1.089	1.682	0.0066*
English (β_{10})	1.602	1.189	2.185	0.0023*
IT (β_{11})	1.232	0.871	1.758	0.2430
Communication (β_{12})	1.015	0.760	1.368	0.9190
Teamwork (β_{13})	0.773	0.541	1.099	0.1527
Selfdevelopment (β_{14})	0.817	0.595	1.126	0.2000

*Note: * Significant with alpha 0.05*

3.2.3. Model significance and model fit

Goodness of fit can be evaluated through complementary statistics, with the deviance test providing a global assessment of model and data agreement. Meanwhile the Hosmer–Lemeshow (HL) test offering a grouped comparison of observed and predicted probabilities. In Table 4, the residual deviance is 737.07 with 618 degrees of freedom, producing a p-value of 0.00066, a result that indicates a statistically significant discrepancy between the fitted model and the saturated model and therefore suggests imperfect overall fit. However, the HL test yields $X^2 = 15.452$, df = 8, and a p-value of 0.05092, which is marginally above the conventional 0.05 threshold. This outcome implies that, when the data are evaluated within decile-based risk groups, the predicted and observed frequencies do not differ significantly, and the model demonstrates acceptable calibration at this level of aggregation. Taken together, these results indicate that although the global deviance measure flags some degree of misfit, the HL test provides evidence that the model's predicted probabilities remain reasonably aligned with observed outcomes, supporting cautious interpretation of the GLM estimates. To assess the adequacy of the model, the dispersion parameter was estimated as the ratio of residual deviance to the degrees of freedom, yielding a value of approximately 1.19. A dispersion parameter close to 1 indicates that the variability in the data is consistent with the assumed model structure [20]. Since the observed ratio is proximate to unity and remains below the practical threshold of 1.5, the model is deemed to exhibit a satisfactory fit with only mild overdispersion [21].

Table 4. Result of Goodness of fit test

Goodness of Fit Test	Statistics	df	p-value
Deviance Test	Deviance = 737.07	618	0.00066
Hosmer–Lemeshow Test	X^2 = 15.452	8	0.05092

Goodness of fit can also be assessed by examining the model's discriminative ability through the Receiver Operating Characteristic (ROC) curve. ROC analysis provides a model-based evaluation of how well the classifier distinguishes between outcome categories across a continuum of decision thresholds. In this framework, the ROC curve plots the true positive rate against the false positive rate, offering a threshold-independent view of performance, while the area under the curve (AUC) summarizes the model's overall capacity to separate individuals with and without the outcome. This assessment complements calibration-oriented goodness-of-fit tests by focusing specifically on predictive discrimination, thereby contributing to a more comprehensive evaluation of model adequacy.

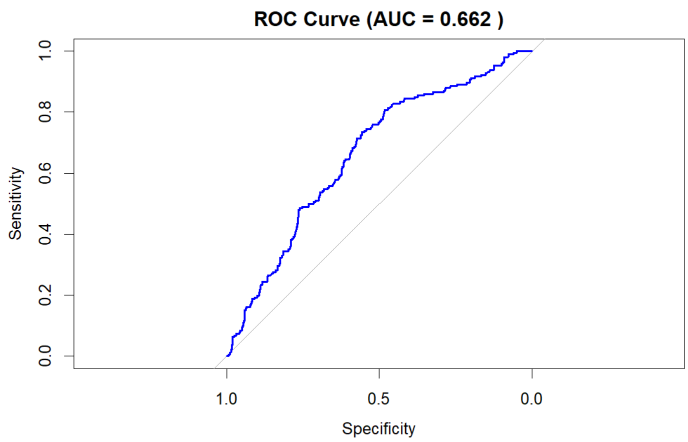

Figure 3. Receiver Operating Characteristic (ROC) Curve from Model

The ROC curve provides an evaluation of the model's discriminative capacity by plotting sensitivity against 1 for specificity across a continuum of probability thresholds. The estimated area under the curve (AUC) is 0.662, indicating that the model achieves a modest ability to distinguish between individuals with income above and below the designated threshold. In practical terms, an AUC of this magnitude suggests that, when presented with a randomly selected pair of respondents which is one with a positive outcome and one with a negative outcome then the model will assign a higher predicted probability to the positive case approximately 66% of the time. Although this level of discrimination exceeds that of random classification, it does not reach the levels typically associated with strong predictive performance, implying that the model captures some, but not all, of the systematic variation relevant to the outcome.

3.3. Discussion

The binary GLM results reveal that several predictors exert significant and theoretically coherent effects on the probability of graduates attaining income above the UMR threshold, with odds ratios demonstrating that self-reported technical competence and English proficiency positively influence earnings. Each unit increase in these competencies corresponds to a meaningful multiplicative rise in the odds of high-income attainment, reinforcing human-capital theory which posits that skill accumulation translates into labor-market returns. Empirical evidence from recent international studies corroborates this pattern. Skill-based indicators were among the strongest determinants of alumni income in a large-scale machine-learning model [22]. Similarly, Henshaw et al. demonstrated that advanced competencies contributed significantly to wage differentiation in UK graduate datasets [23]. Additional literature underscores the economic value of English proficiency, with language skills consistently associated with wage premiums across global labor markets [24], [25] ,[26] . These findings collectively support the interpretation that competencies remain central determinants in shaping early-career economic outcomes.

The most prominent predictor in the model is tuition-financing. Graduates supported by scholarships demonstrate substantially higher odds of achieving income above UMR relative to self-funded peers. This effect is consistent with studies showing that scholarship recipients benefit from both selection advantages and structural supports that enhance employability. Indonesian evidence from the LPDP scholarship program indicates elevated economic contributions among scholarship alumni [27]. International research similarly documents that financial aid improves academic persistence, reduces dropout risk, and enhances labor-market integration, often translating into higher earnings [28], [29]. A plausible mechanism is that financial support alleviates economic pressures, enabling students to invest more time in internships, extracurricular development, and networking activities repeatedly identified as predictors of improved early-career salaries. For example, Delis demonstrated that structured work-integrated learning experiences significantly improve graduate salary outcomes, especially when financial barriers are reduced [30]. Thus, the strong odds ratio associated with tuition financing is not merely a statistical artifact but reflects deeper institutional, socio-economic, and structural contributions to graduates' labor-market trajectories.

Despite statistically significant predictors, the model exhibits modest discriminative performance (AUC ≈ 0.66). Such modest predictive power aligns with the broader literature, which repeatedly shows that even sophisticated models deployed on extensive graduate datasets rarely achieve high accuracy when limited to pre-graduation characteristics. Wang et al. reported only moderate accuracy in predicting starting salaries using machine-learning models, arguing that firm-level heterogeneity, regional factors, and job-sector dynamics explain substantial additional variance [31]. Li's LSTM-based salary prediction study similarly noted that the majority of predictive power comes from contextual labor-market variables rather than individual competencies [32]. Studies by Zhou and Ramezani further affirm that earnings dynamics depend on multifactorial economic, social, and behavioral determinants beyond educational attributes [33]. This correspondence with international findings suggests that the modest AUC in the present study reflects a common structural limitation inherent to graduate-income modeling.

Measurement constraints likely also contribute to limited predictive accuracy. Self-reported competencies such as communication, teamwork, and self-development are subject to bias and often correlate weakly with actual job performance. Evidence from vocational behavior research shows that employer-rated soft skills predict job outcomes substantially more accurately than self-reports [34], [15]. Additionally, dichotomizing income into above/below UMR reduces informational granularity and may obscure important income variation; numerous methodological studies warn that dichotomization reduces statistical power and prediction accuracy [35], [36]. This suggests that future modeling should consider continuous or ordinal income outcomes to retain more variance. Nevertheless, significant predictors remain institutionally meaningful. Competency development especially in technical and linguistic domains. These aligns with findings from higher-education research advocating competency-based curricula as critical to boosting employability [37], [38]. The strong effect of scholarship funding reinforces the value of expanding financial aid as a tool for increasing equity and labor-market success. At the same time, the modest AUC underscores that pre-graduation data alone cannot fully capture earnings complexity, pointing to the need for integrated graduate tracking systems. International evidence shows that combining alumni survey data with labor-market indicators, employer data, and regional economic measures substantially improves model accuracy [39], [40]. Institutions should therefore consider longitudinal approaches and richer data infrastructures to improve predictive modeling and enhance strategic decision-making.

In summary, the significant odds ratios for academic competence, English proficiency, and study duration highlight the importance of human-capital attributes in shaping graduate income outcomes. Graduates with stronger academic competencies and English proficiency exhibit higher odds of earning above the regional minimum wage, while longer study duration is also associated with increased income likelihood. Yet the model's limited discrimination confirms that income formation is multifactorial and dependent on labor-market dynamics and unobserved personal factors. Enhancing competency development, particularly in academic and language-related domains, together with improving data integration, represent promising institutional avenues for strengthening graduate employability and economic mobility.

4. CONCLUSION

The results of this study demonstrate that specific pre-graduation characteristics particularly an academics knowledge, English proficiency, and tuition-financing such as scholarship exert meaningful influence on the likelihood of alumni earning above the UMR threshold. The significant odds ratios associated with these predictors confirm that both individual human-capital attributes and institutional support mechanisms contribute to early-career income differentiation. Scholarship-funded graduates exhibit substantially higher odds of attaining higher income, suggesting that financial aid not only reduces economic barriers during study but may also enable greater access to skill-building opportunities that translate into improved labor-market outcomes. Competency-related predictors similarly underscore the continued relevance of skill formation in shaping economic opportunities in increasingly competitive and globalized labor markets. Despite these associations, the model's modest discriminative performance indicates that pre-graduation variables alone cannot fully capture the complexity of income formation. This study therefore contributes to the broader body of knowledge by illustrating both the strengths and limitations of competency- and financing-based predictors within a GLM framework, reinforcing findings from recent international research that highlight the multifactorial nature of graduate earnings. The results emphasize the importance of integrating richer labor-market data, employer evaluations, and longitudinal career trajectories in future predictive modeling. Further research might also explore non-linear modeling approaches or continuous income measures to enhance explanatory and predictive power. Overall, the findings support institutional policies that expand scholarship access and strengthen competency development while encouraging more holistic, data-enhanced approaches to understanding graduate economic outcomes.

Funding Information

The authors declare that this research received no external funding.

Conflict Of Interest Statement

The authors declare that there is no conflict of interest regarding the publication of this paper.

Informed Consent

The authors confirm that informed consent was secured from all study participants.

Data Availability

The data that support the findings of this study are available on request from the corresponding author. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

[1] T. Hooley, J. Hanson, and L. Clark, "Exploring students' and graduates' attitudes to the process of transition to the labour market," *Industry and Higher Education*, vol. 37, no. 1, pp. 121–134, Feb. 2023, doi: 10.1177/09504222221111298.

[2] J. Robson, "Graduate employability and employment," in *Assessing the Contributions of Higher Education*, Edward Elgar Publishing, 2023, pp. 177–196. doi: 10.4337/9781035307173.00019.

[3] R. Blokker, J. Akkermans, M. Tims, P. Jansen, and S. Khapova, "Building a sustainable start: The role of career competencies, career success, and career shocks in young professionals' employability," *J Vocat Behav*, vol. 112, pp. 172–184, Jun. 2019, doi: 10.1016/j.jvb.2019.02.013.

[4] N. Nacong, Hajar, and Andri, "Pengaruh Nilai Perilaku Dan Nilai Akademik Terhadap Keberhasilan Alumni Saat di Dunia Kerja Menggunakan Regresi Linear Berganda," *Innovative: Journal Of Social Science Research*, vol. 4, no. 1, 2024, Accessed: Dec. 01, 2025. [Online]. Available: https://j-innovative.org/index.php/Innovative/article/view/9022

[5] M. Irfany, "Analisis Faktor-Faktor yang Mempengaruhi Masa," Universitas Islam Indonesia, Jogjakarta, 2007.

[6] S.-I. Lee, H. Lee, P. Abbeel, and A. Y. Ng, "Efficient L 1 Regularized Logistic Regression," in *In Proceedings of the 21st national conference on Artificial intelligence-Volume 1*, 2006. [Online]. Available: www.aaai.org

[7] D. Straszak and N. K. Vishnoi, "IRLS and Slime Mold: Equivalence and Convergence," Jan. 2016, [Online]. Available: http://arxiv.org/abs/1601.02712

[8] L. Fahrmeir and G. Tutz, *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2nd ed. New York, NY: Springer New York, 2001. doi: 10.1007/978-1-4757-3454-6.

[9] R. Finnie and M. Frenette, "Earning differences by major field of study: evidence from three cohorts of recent Canadian graduates," *Econ Educ Rev*, vol. 22, no. 2, pp. 179–192, Apr. 2003, doi: 10.1016/S0272-7757(02)00003-1.

[10] M. Francesconi and M. Parey, "Early gender gaps among university graduates," *Eur Econ Rev*, vol. 109, pp. 63–82, Oct. 2018, doi: 10.1016/j.euroecorev.2018.02.004.

[11] 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.
[12] T. T. A. Ngo, "The Importance of Soft Skills for Academic Performance and Career Development—From the Perspective of University Students," *International Journal of Engineering Pedagogy (iJEP)*, vol. 14, no. 3, pp. 53–68, Apr. 2024, doi: 10.3991/ijep.v14i3.45425.
[13] M. V. Gruzdev, I. V. Kuznetsova, I. Y. Tarkhanova, and E. I. Kazakova, "University graduates' soft skills: The employers' opinion," *European Journal of Contemporary Education*, vol. 7, no. 4, pp. 690–698, 2018, doi: 10.13187/ejced.2018.4.690.
[14] J. Gleeson, R. Black, A. Keddie, and C. Charles, "Graduate capitals and employability: Insights from an Australian university co-curricular scholarship program," *Pedagogy, Culture & Society*, vol. 32, no. 2, pp. 341–359, Mar. 2024, doi: 10.1080/14681366.2022.2038251.
[15] C. Succi and M. Canovi, "Soft skills to enhance graduate employability: comparing students and employers' perceptions," *Studies in Higher Education*, vol. 45, no. 9, pp. 1834–1847, Sep. 2020, doi: 10.1080/03075079.2019.1585420.
[16] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. Wiley, 2013. doi: 10.1002/9781118548387.
[17] S. Pérez-Fernández, P. Martínez-Camblor, P. Filzmoser, and N. Corral, "Visualizing the decision rules behind the ROC curves: understanding the classification process," *AStA Advances in Statistical Analysis*, vol. 105, no. 1, pp. 135–161, Mar. 2021, doi: 10.1007/s10182-020-00385-2.
[18] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, Oct. 2018, doi: 10.1016/j.neunet.2018.07.011.
[19] P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein, "A simulation study of the number of events per variable in logistic regression analysis," *J Clin Epidemiol*, vol. 49, no. 12, pp. 1373–1379, Dec. 1996, doi: 10.1016/S0895-4356(96)00236-3.
[20] P. McCullagh and J. A. Nelder, *Generalized Linear Models*. Routledge, 2019. doi: 10.1201/9780203753736.
[21] A. F. Zuur, E. N. Ieno, N. Walker, A. A. Saveliev, and G. M. Smith, *Mixed effects models and extensions in ecology with R*. New York, NY: Springer New York, 2009. doi: 10.1007/978-0-387-87458-6.
[22] D. A. Gomez-Cravioto, R. E. Diaz-Ramos, N. Hernandez-Gress, J. L. Preciado, and H. G. Ceballos, "Supervised machine learning predictive analytics for alumni income," *J Big Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00559-6.
[23] B. Henshaw, B. K. Mishra, W. Sayers, and Z. Pervez, "Unveiling the Impact of Socioeconomic and Demographic Factors on Graduate Salaries: A Machine Learning Explanatory Analytical Approach Using Higher Education Statistical Agency Data," *Analytics*, vol. 4, no. 1, Mar. 2025, doi: 10.3390/analytics4010010.
[24] O. Galvez-Soriano, "Foreign language skills and labor market outcomes: The case of English in Mexico," *Econ Educ Rev*, vol. 108, p. 102702, Oct. 2025, doi: 10.1016/j.econedurev.2025.102702.
[25] R. Godoy *et al.*, "Language skills and earnings: Evidence from a pre-industrial economy in the Bolivian Amazon," *Econ Educ Rev*, vol. 26, no. 3, pp. 349–360, Jun. 2007, doi: 10.1016/j.econedurev.2005.12.001.
[26] C. Dustmann and F. Fabbri, "Language Proficiency And Labour Market Performance Of Immigrants In The Uk*."
[27] R. E. B. Gufron, M. Simanjuntak, and T. Novianti, "Program Beasiswa dan Peningkatan Kinerja Dampak Kepemimpinan, Kontribusi Sosial dan Ekonomi Alumni LPDP," *Jurnal Aplikasi Bisnis dan Manajemen*, Sep. 2022, doi: 10.17358/jabm.8.3.926.
[28] J. Mountjoy, "Community Colleges and Upward Mobility," *American Economic Review*, vol. 112, no. 8, pp. 2580–2630, Aug. 2022, doi: 10.1257/aer.20181756.
[29] G. L. Bernal, L. K. Abadia, L. E. Alvarez-Arando, K. De Witte, L. K. Abadía, and L. E. Álvarez-Arando, "Financial Aid Uncertainty and Low-Income Students' Higher Education

Preferences Financial Aid Uncertainty and Low-Income Students' Higher Education Preferences 1," 2023.
[30] A. Delis and C. Jones, "The impact of work placements on graduate earnings," *Studies in Higher Education*, vol. 48, no. 11, pp. 1708–1723, Nov. 2023, doi: 10.1080/03075079.2023.2211999.
[31] P. Wang, W. Liao, Z. Zhao, and F. Miu, "Prediction of Factors Influencing the Starting Salary of College Graduates Based on Machine Learning," *Wirel Commun Mob Comput*, vol. 2022, no. 1, Jan. 2022, doi: 10.1155/2022/7845545.
[32] F. Li, N. A. Majid, and S. Ding, "Unlocking the potential of LSTM for accurate salary prediction with MLE, Jeffreys prior, and advanced risk functions," *PeerJ Comput Sci*, vol. 10, p. e1875, Feb. 2024, doi: 10.7717/peerj-cs.1875.
[33] M. Zhou and R. Ramezani, "A Deep Dive into the Factors Influencing Financial Success: A Machine Learning Approach."
[34] A. Payne, "Assessing Post-Secondary Graduates' Soft Skills, Job Search and Employment Outcomes," University of Guelph, 2018.
[35] N. S. Al-Nassar, A. A. Alhajjaj, and A. Bleady, "Potential predictors of length of study for finance and accounting degrees: evidence from a public business school in Saudi Arabia," *Heliyon*, vol. 8, no. 6, p. e09636, Jun. 2022, doi: 10.1016/j.heliyon.2022.e09636.
[36] J. E. Hunter and F. L. Schmidt, "Dichotomization of continuous variables: The implications for meta-analysis.," *Journal of Applied Psychology*, vol. 75, no. 3, pp. 334–349, Jun. 1990, doi: 10.1037/0021-9010.75.3.334.
[37] Md. H. Rahman, Md. M. Islam, M. Parvin, N. Sultana, and Md. A. Rahman, "Institutional Assistance and Adaptation of Fresh Graduates Facing Employment Crisis and Poverty," *Journal of Business Studies*, vol. 03, no. 01, pp. 119–143, 2022, doi: 10.58753/jbspust.3.1.2022.8.
[38] A. M. Al Shukaili, K. Al Kindi, N. M. Kassim, Z. Ahmed, and K. Al Hosni, "Can government financial support enhance job creations: insights from Oman," *Journal of Science and Technology Policy Management*, vol. 14, no. 5, pp. 807–829, Aug. 2023, doi: 10.1108/JSTPM-07-2021-0100.
[39] G. Groves and I. Palmer, "Tracking Graduates into the Workforce: Connecting Education and Labor Market," 2015. [Online]. Available: www.nga.org.
[40] F. R. Jones, M. Mardis, C. McClure, and E. Randeree, "Alumni tracking: Promising practices for collecting, analyzing, and reporting employment data," *Journal of Higher Education Management*, vol. 32, no. 1, 2017.