APPLICATION OF FIRTH'S LOGISTIC REGRESSION IN ANALYZING EMPLOYMENT STATUS OF EDUCATED WOMEN IN RIAU

Kayla Azka Dhiya Tsabithah[1)]
Liza Kurnia Sari[2)]
[1,2)] Politeknik Statistika STIS, Jakarta, Indonesia
e-mail: kayla.dhiya21@gmail.com

ABSTRACT

The ongoing demographic bonus in Indonesia provides a strategic opportunity to promote inclusive development. To optimize this potential, the involvement of educated women is crucial and not just rely on the contribution of men. However, the women's Labor Force Participation Rate (LFPR) has remained stagnant over the past two decades and gender inequality in employment persists to this day. Riau Province requires the most attention, as in the last three years it has recorded the lowest women's LFPR nationally. Interestingly, data show that the proportion of educated women who are employed is actually lower than that of less educated women. This study applies Firth's logistic regression to analyse the employment status of educated women in Riau Province in 2024. The method is employed to reduce estimation bias that may arise from data imbalance. The results indicate that ICT use, age, marital status, and place of residence have significant effects. These findings can support the development of more effective strategies to increase the employment status of educated women, particularly by strengthening ICT access and considering sociodemographic characteristics.

Keywords: educated women, employment status, Firth's logistic regression.

INTRODUCTION

Currently, Indonesia is experiencing a demographic bonus, a condition in which the number of people of productive age (15–64 years) exceeds that of the nonproductive population. This condition can drive economic growth, particularly in the labor sector (Adriani & Yustini, 2021). The benefits of this demographic bonus are more likely to be realized under ideal conditions where men and women play equally significant roles in all fields, including the labor market. As stated by the World Bank (2011), gender equality can enhance productivity and foster national development.

Unfortunately, employment in Indonesia remains marked by gender inequality. The government has made efforts to establish a gender responsive legal framework for labor, as stated in Law Number 13 of 2003, Article 5, which ensures that every worker has equal opportunities without discrimination in obtaining employment. This implies that both men and women have equal rights and opportunities. Therefore, continuous efforts are needed to address this issue in order to achieve gender equality, in line with the fifth goal of the Sustainable Development Goals (SDGs).

An important indicator in measuring labor conditions is the Labor Force Participation Rate (LFPR). The LFPR reflects the level of involvement of the economically active population in a region (Koto et al., 2025). As shown in Figure 1, there remains a gap between the LFPR of men and women. Moreover, the progress of women's LFPR in Indonesia has been relatively slow. Over the past two decades, the women's LFPR in Indonesia has remained stagnant at around 50 percent, while the men's LFPR has consistently stayed at about 80 percent. To this day, the women's LFPR remains lower than men's in all provinces. Particular attention should be given to Riau Province, which has recorded the lowest women's LFPR in Indonesia from 2022 to 2024 (BPS, 2022, 2023, 2024b). This poses a challenge for the government to achieve both the gender equality goal of

SDGs and the target women's LFPR of 70 percent by 2045, as stated in Law No. 59 of 2024 on the National Long Term Development Plan 2025-2045.

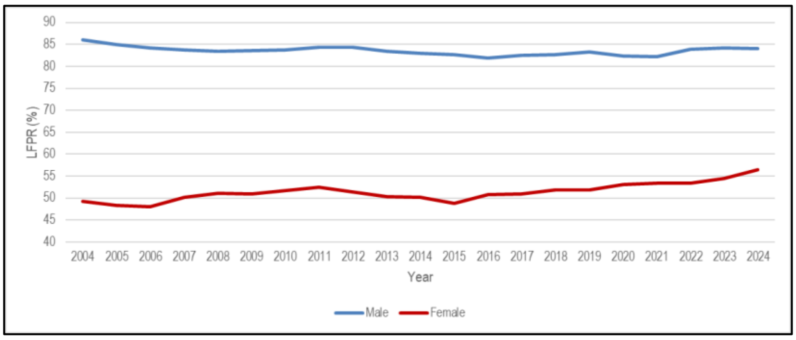

Figure 1. LFPR by Gender in Indonesia 2004-2024
Source: BPS

The labor conditions of women become more complex when associated with the group of educated women, namely those who have completed at least senior high school education. Despite having educational qualifications, educated women still face their own barriers in obtaining employment opportunities. In relation to this, according to human capital theory, education is considered a form of investment in human capital that directly affects the improvement of employment opportunities and labor productivity (Mankiw & Harris, 1998). Education is believed to enhance individual skills and competencies, thereby providing greater opportunities to participate in the labor market.

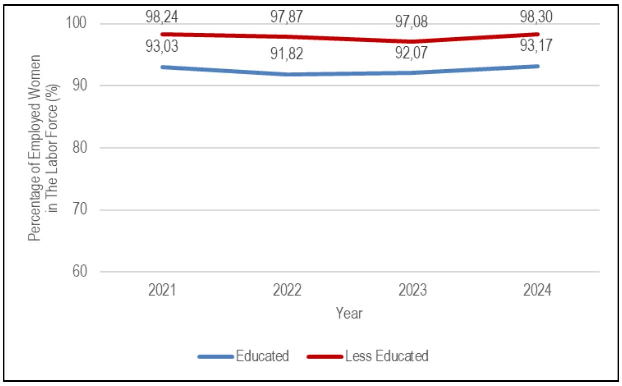

Figure 2. Percentage of Employed Women in The Labor Force in Riau Province by Education Level 2021-2024
Source : BPS; processed

However, data from BPS show that the percentage of educated women in the labor force who are employed in Indonesia, including in Riau Province, has consistently been lower than that of less educated women over the years. As shown in Figure 2, the percentage of less educated women in the labor force who are employed in Riau Province remains consistently around 97 to 98 percent. Meanwhile, for educated women, the percentage is only around 91 to 93 percent. This situation persists even though Riau Province has considerable potential. This is reflected in educational and ICT attainment. According to BPS, the percentage of educated women in Riau is always above the national average, indicates that Riau has relatively good female human resources. Then, its ICT Development Index ranks among the ten highest in Indonesia (BPS, 2024a). According to Simanjuntak (1998), as cited in Shantika (2019), the higher the level of education attained, the greater the opportunity to obtain employment. Then, Kusumawardhani et al. (2023) explain that ICT tends to generate more favorable labor market outcomes for more educated individuals. The persistence of this gap suggests that such potential has not yet been fully converted into employment opportunities for educated women. This issue also indicates that the labor market's capacity to provide jobs that match the qualifications of educated women remains limited.

This issue requires appropriate action, as its continuation could lead to inefficiency in educational investment. Moreover, it may also hinder economic development and the realization of the demographic bonus potential. As stated by Okorie (2013) in Shantika (2019), women's education plays an important role in accelerating economic development and reducing poverty, thereby improving societal welfare. Building on this, when more women participate in the workforce and become economically empowered, it can generate a multiplier effect on economic growth and development across various sector (Damayanti, 2021). Therefore, women with the competencies they possess should be given equal opportunities to actively contribute to economic development and improve overall wellbeing (Kurniasih et al., 2022).

Studies by Ngoa & Song (2021) in Africa and Suhaib & Kartiasih (2024) in Papua Province using panel regression find that ICT (internet access, mobile phone use, and computer use) have a positive effect on women's labor force participation. In addition, study by Lusiyanti & Wicaksono (2020) in Indonesia, employing a probit regression shows that sociodemographic characteristics, including age, marital status, and place of residence also influence women's employment status. However, existing studies focusing on Riau Province remains scarce and the application of the Firth's Logistic Regression method in studies of educated women's employment is still limited. Thus, this study aims to apply Firth's logistic regression model in identify the variables that influence the employment status of educated women in Riau Province 2024. Parameter estimation is conducted using Penalized Maximum Likelihood Estimation (PMLE) approach. This approach is applied due to the imbalance in the employment participation data of educated women. According to BPS (2024c), the proportion of employed educated women in the labor force is 93.17 percent, while the remaining 6.83 percent are unemployed. Although MLE can be technically applied, PMLE is more suitable for handling imbalanced data. Under such condition, MLE tends to produce biased parameter estimates that deviate from the true parameter values. Therefore, PMLE is used to reduce this bias so that the resulting estimates are closer to the true parameter values on average (Firth, 1993).

METHOD

Scope of Research

The data used in this study were obtained from BPS, specifically from the raw data of the August 2024 Sakernas (National Labor Force Survey). The unit of analysis in this study is women in the labor force aged 15-64 years with a minimum education level of senior high school or equivalent. After the selection process to determine the unit of analysis, a total of 2,316 samples were obtained.

The variables used in this study consist of a response variable and predictor variables. The response variable is the employment status of educated women, which is categorized into two groups: employed and unemployed. Meanwhile, the predictor variables are ICT use, age, marital status, place of residence, and training. In this study training refers to participation in activities that provide specific skills conducted by either government or private institutions including training carried out in the workplace or related to other activities undertaken by the respondents. This definition based on BPS guidelines.

The predictor variables in this study are selected based on concept of employability by McQuaid & Lindsay (2005) which explains that several aspects influence whether individuals are employed, including individuals' basic ICT skills, demographic characteristics, and social conditions. The selection of these variables is further supported by previous studies such as Suhaib & Kartiasih (2024) who finds that ICT influences women's labor force participation and studies by Lusiyanti & Wicaksono (2020) which show that sociodemographic characteristics also have an influence. The operational definitions of the variables used in this study are presented in Table 1 below.

Table 1. Operational definition of variables

Variables	Description	Category	Code
Y	Employment Status	Employed	1
		Unemployed*	0
X_1	ICT Use	Using 3 Devices (using all 3 : hp, internet, and computer)	1
		Using <3 Devices*	0
X_2	Age	Not Young Age	1
		Young Age (15-24)*	0
X_3	Marital Status	Ever Married	1
		Unmarried*	0
X_4	Place of Residence	Rural	1
		Urban*	0
X_5	Training	Ever	1
		Never*	0

Note: *Reference Category

Method of Analysis

This study uses two analytical methods: descriptive analysis and inferential analysis. The descriptive analysis is conducted using tables to provide an overview of the characteristics of educated women based on their employment status and the variables presumed to influence it. Meanwhile, the inferential analysis is carried out using Firth's logistic regression method to identify the significant variables and their tendencies in influencing the employment status of educated women. The stages of the inferential analysis include parameter estimation with PMLE, simultaneous significance parameter test, partial significance parameter test, model fit test, checking classification accuracy measures, and model interpretation.

1. Binary Logistic Regression

The regression method is commonly used in data analysis to describe the relationship between a response variable and one or more predictor variables. When the response variable is categorical, the most widely used approach is logistic regression (Hosmer et al., 2013). There are several types of logistic regression, one of which is binary logistic regression, applied when the response variable is dichotomous or consists of two categories.

In logistic regression with p predictor variables, (Hosmer et al., 2013) define the model as follows:

$$\pi(\boldsymbol{x}) = \frac{\mathrm{e}^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}}{1+\mathrm{e}^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}} \tag{1}$$

In equation (1) the relationship between $\pi(\boldsymbol{x})$ and its parameters is not linear, so a logit transformation is performed to make it linear. Below are the results of the logit transformation that produce the equation for the model used in this study:

$$g(\boldsymbol{x}) = \ln\left[\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right] = g(\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p Y_i = \sum(Benefit) - \sum(Cost) \quad (2)$$

Keterangan:
$g(\boldsymbol{x})$ = logit function
$\pi(\boldsymbol{x})$ = chance of success of an event
β_j = coefficient of the j-th parameter with $j = 1, 2, \dots, p; p = 5$
x_j = value of the j-th parameter with $j = 1, 2, \dots, p; p = 5$

In estimating its parameters, binary logistic regression uses the Maximum Likelihood Estimation (MLE) method (Hosmer et al., 2013). To apply MLE, a likelihood function must first be constructed to represent the probability of the observed data given the unknown parameter values. MLE then seeks the parameter values that maximize the likelihood function.

2. Firth's Logistic Regression

Firth's logistic regression is an alternative form of binary logistic regression designed to handle data imbalance problem. In binary logistic regression, an imbalance between the two classes may occur. Data imbalance can lead to biased parameter estimates when using the MLE method (Amini & Oktora, 2021).

Firth (1993) proposed a parameter estimation method that reduces MLE bias by modifying the MLE score function, which is then referred to as the Penalized Maximum Likelihood Estimation (PMLE). The score function is the first derivative of the loglikelihood function with respect to the parameters. The MLE estimates are obtained by finding the solution to:

$$L(\boldsymbol{\beta}) = U(\boldsymbol{\beta}) = 0 \quad (3)$$

where $U(\boldsymbol{\beta})$ is the score function. The score function is modified to reduce bias in parameter estimation by adding a Fisher information-based penalty as follows:

$$U^*(\boldsymbol{\beta}) = U(\boldsymbol{\beta}) - \mathrm{I}(\boldsymbol{\beta})\mathrm{B}(\boldsymbol{\beta}) \quad (4)$$

Keterangan:
$U^*(\boldsymbol{\beta})$ = PMLE score function
$U(\boldsymbol{\beta})$ = MLE score function
$\mathrm{I}(\boldsymbol{\beta})$ = fisher information matrix
$\mathrm{B}(\boldsymbol{\beta})$ = asymptotic bias of the MLE estimate

The solution of $U^*(\boldsymbol{\beta})$ yields a new estimate $\boldsymbol{\beta}^*$, which is closer to the true parameter value or less biased. The PMLE function proposed by Firth (1993) is expressed as follows:

$$l^*(\boldsymbol{\beta}) = l(\boldsymbol{\beta})|I(\boldsymbol{\beta})|^{1/2} \quad (5)$$

Using the loglikelihood function, the equation can be written as follows:

$$L^*(\boldsymbol{\beta}) = \ln[l^*(\boldsymbol{\beta})] = \ln[l(\boldsymbol{\beta})] + \frac{1}{2}\ln[|I(\boldsymbol{\beta})|] \qquad (6)$$

Keterangan:
$l(\boldsymbol{\beta})$ = likelihood function of MLE
$l^*(\boldsymbol{\beta})$ = penalized likelihood function of PMLE
$L^*(\boldsymbol{\beta})$ = penalized loglikelihood function of PMLE
$|I(\boldsymbol{\beta})|$ = jeffrey's invariant prior

3. Simultaneous Parameter Significance Test

The simultaneous parameter significance test is conducted to evaluate whether at least one predictor variable has a significant effect on the model. If the predicted values obtained using the variables in the model are better or more accurate than those obtained without the variables, the corresponding variable or parameter is considered significant (Hosmer et al., 2013). The test is performed using the Likelihood Ratio Test (LRT) with the test statistic G. The hypothesis and test statistic for the LRT are as follows:
Hypothesis:
$H_0 : \beta_1 = \beta_2 = \cdots = \beta_5 = 0$
H_1: there is at least one $\beta_j \neq 0, j = 1, 2, \ldots, 5$

Test Statistic:

$$G = -2\, ln\left(\frac{l_0}{l_1}\right) \sim \chi^2_{(5)} \qquad (7)$$

H_0 is rejected if $G > \chi^2_{(0.05;5)}$ = 11.070 or when the p-value < 0.05, indicating that at a 5 percent significance level, there is at least one predictor variable that significantly affects the employment status of educated women.

4. Partial Parameter Significance Test

The partial parameter significance test is conducted to determine the effect of each predictor variable on the response variable. The test is performed using the Wald Test with the test statistic W_j. The hypothesis and test statistic are as follows:
Hypothesis:
$H_0 : \beta_j = 0, j = 1, 2, \ldots, 5$
$H_1 : \beta_j \neq 0, j = 1, 2, \ldots, 5$

Test Statistic:

$$W_j = \frac{\widehat{\beta}_j}{\widehat{SE}(\widehat{\beta}_j)} \sim \mathrm{N}_{(0,1)} \qquad (8)$$

H_0 is rejected if $|W_j| > Z_{(0.05/2)}$ = 1.96 or when the p-value < 0.05, indicating that at the 5% significance level, the j-th predictor variable has a significant effect on the employment status of educated women.

5. Model Fit Test

The model fit test is conducted using the Hosmer and Lemeshow Test. This test aims to evaluate how well the predicted probabilities from the logistic model correspond to the actual observed values. In this case, the model is considered a good fit if the predicted probabilities do not differ significantly from the observed data. The hypothesis and test statistic are as follows:

Hypothesis:

H_0 : model fits the data

H_1 : model doesn't fit the data

Test Statistic:

$$\hat{C} = \sum_{k=1}^{g} \frac{(o_{1k} - n'\bar{\pi}_k)^2}{n'\bar{\pi}_k(1 - \bar{\pi}_k)} \sim \chi^2_{(6)} \tag{9}$$

H_0 fails to be rejected if $\hat{C} < \chi^2_{(0.05;6)}$ = 12.592 or if the p-value > 0.05, indicating that the constructed model fits the data used in the analysis.

6. Classification Accuracy Measures

Classification Table

Classification table is one of classification accuracy measure. Classification table is the result of a cross tabulation between the response variable (y) and a dichotomous variable whose values are derived from the estimated logistic probabilities (Hosmer et al., 2013). In its application, a cutoff point (c), must be determined and each estimated probability is compared with c. A commonly used cutoff value is 0.5. However, the use of a 0.5 cutoff does not provide good accuracy for imbalanced data. This is because it is sensitive to the relative sizes of the two component groups and tends to classify observations into the larger group (Hosmer et al., 2013). Therefore, it is necessary to identify an optimal cutoff value.

The optimal cutoff value can be obtained using the Youden Index, which has been broadly applied in practice (Yin & Tian, 2014). It is independent of the relative sizes of the component groups. The optimal cutoff is determined by comparing the Youden Index (J) values across all candidate cutoffs. Where the cutoff with the largest value of $J = sensitivity + specificity - 1$ is selected as the optimal cutoff (Youden, 1950).

Sensitivity is a measure of the proportion of true positives that are correctly classified, while specificity is a measure of the proportion of true negatives that are correctly classified. In addition, accuracy is a measure of the overall proportion of correctly classified, including both true positives and true negatives, relative to the total number of observations. The classification table is presented below, as shown in Table 2.

Table 2. Classification Table

Predicted	**Observed**	
	$y = 1$	$y = 0$
$y = 1$	True Positive (TP)	False Positive (FP)
$y = 0$	False Negative (FN)	True Negative (TN)
Total	P = TP+FN	N = FP+TN

Source: Hosmer et al. (2013)

Receiver Operating Characteristic (ROC)

A more comprehensive measure of classification accuracy is the area under ROC curve. The ROC curve illustrates the relationship between sensitivity and 1–specificity across all possible cutoff values. This measure has been widely adopted as a standard for evaluating model performance. The area under the ROC curve ranges from 0.5 to 1.0 and reflects the model's ability to distinguish between the two categories of the response variable, with values closer to 1 indicating better discriminative performance (Hosmer et al., 2013).

7. Model Interpretation

In conducting the interpretation, the first step is to determine the link function or the function of the response variable that produces a linear function of the predictor variables (Hosmer et al., 2013). In logistic regression, the link function is the logit (log-odds) transformation. To simplify the interpretation on the log-odds scale, it is common to convert it into an odds ratio, referred to as OR (Niu, 2020). The results of conversion produce an OR estimation equation as shown below:

$$\widehat{OR} = e^{\hat{\beta}_j} \quad (10)$$

The value of $\widehat{OR}$ is used for interpreting the model. In this study, the response variable is the employment status of educated women, which consists of two categories: employed (y = 1) and unemployed (y = 0). For categorical predictor variables, this value is used to indicate the tendency of educated women being employed in a particular category of a predictor variable compared to its reference category.

RESULTS AND DISCUSSION

Descriptive Analysis

Table 3 presents a general overview of the characteristics of educated women, also viewed based on employment status and each predictor variable presumed to influence it.

Table 3. Characteristics of Educated Women Based on Employment Status and Predictor Variables in Riau 2024

Variabel	**Category**	**Percentage (%)**	**Employment Status (%)**	
			Employed	**Unemployed**
ICT Use	Using 3 devices	40.73	**95.64**	4.36
	Using <3 devices	59.27	91.45	8.55
Age	Not Young Age	74.36	**96.9**	3.1
	Young Age	25.64	82.29	17.71

Variabel	Category	Percentage (%)	Employment Status (%)	
			Employed	Unemployed
Marital Status	Ever Married	65.15	**97.84**	2.16
	Unmarried	34.85	84.4	15.6
Place of Residence	Rural	45.55	**94.18**	5.82
	Urban	54.45	92.3	7.7
Training	Ever	5.77	**96.95**	3.05
	Never	94.23	92.92	7.08

It is known that most educated women use fewer than three ICT devices, are not in the young age group, ever married, live in urban areas, and never attended training. Furthermore, as shown in Table 3, the proportion of employed educated women is higher among those who use three ICT devices (95.64 percent) compared to those who use fewer than three devices (91.45 percent). This difference indicates that ICT proficiency may support their involvement in employment. A study by Ngoa & Song (2021) stated that ICT enables women to access various types of information more easily, including employment information. In terms of age, the proportion of employed educated women is higher among the non-young age group (96.9 percent). According to human capital theory, as individuals grow older, their accumulation of human capital increases, allowing them to be more competitive in the labor market (Mankiw & Harris, 1998). Moreover, the proportion of employed educated women is higher among those who are ever married (97.84 percent). This may be because ever married women tend to have greater financial responsibilities (Damayanti, 2021). Especially those who become the primary breadwinners after divorce or the loss of a spouse. The proportion of employed educated women is also greater among those living in rural areas (94.18 percent) compared to those living in urban areas. Although access to education is more readily available in cities, urban areas are also characterized by higher job competition (Tsaniyah & Sugiharti, 2021). Training plays a role in enhancing the skills and competencies required in the labor market (Mulugeta, 2021). As shown in Table 3, educated women who have participated in training have a higher employment proportion (96.95 percent) than those who have never attended any training.

Inferential Analysis

1. Parameter Estimation

The parameter estimation results using PMLE are presented in Table 5. Based on these results, the equation of Firth's logistic regression model obtained is as follows:

$$\hat{g}(\boldsymbol{x}) = 0.944 + 0.834X_1 + 0.675X_2 + 1.738X_3 + 0.523X_4 + 0.347X_5 \quad (11)$$

2. Simultaneous Parameter Significance Test

The results of the simultaneous parameter significance test using LRT are presented in the table below.

Table 4. Result of the simultaneous parameter significance test

G	$\chi^2_{(0.05;5)}$	*p-value*
175.700	11.070	0.000

Table 4 shows that the test statistic value G is 175.700 with a p-value = 0.000. This value is greater than $\chi^2_{(0,05;5)}$ = 11.070 and the p-value is smaller than 0.05. This leads to the decision to reject H_0, indicating that at a 5 percent significance level, at least one predictor variable has a significant effect on the employment status of educated women.

3. Partial Parameter Significance Test

After confirming that at least one predictor variable is significant, a partial parameter significance test was performed using the Wald Test.

Table 5. Result of the partial parameter significance test

Variables	$\widehat{\beta}$	$\widehat{SE}\widehat{\beta}$	W	*p-value*	$\widehat{OR}$
Intercept	0.944	0.156	6.036	0.000	2.570
ICT Use (X_1)	0.834	0.202	4.134	0.000	2.301
Age (X_2)	0.675	0.249	2.703	0.007	1.964
Marital Status (X_3)	1.738	0.277	6.277	0.000	5.683
Place of Residence (X_4)	0.523	0.187	2.797	0.005	1.688
Training (X_5)	0.347	0.574	0.605	0.545	1.415

The results of the partial significance test for each variable are presented in Table 5. If the value of test statistic $|W|$ for a predictor variable is greater than $Z_{(0,05/2)}$ = 1.96 or the p-value is less than 0.05, the decision is to reject H_0. Thus, it can be concluded that at a 5 percent significance level, the corresponding predictor variable significantly affects the employment status of educated women. Based on the results, the significant predictor variables are ICT use, age, marital status, and residential area.

4. Model Fit Test

The results of the model fit test using the Hosmer Lemeshow Test are presented in the table below.

Table 6. Result of the model fit test

$\widehat{C}$	$\chi^2_{(0.05;6)}$	*p-value*
4.668	12.592	0.587

As shown in Table 6, the calculated test statistic value $\hat{C}$ is 4.668 which is smaller than $\chi^2_{(0,05;6)}$ = 12.592. Furthermore, the p-value is greater than 0.05, specifically 0.587. Therefore, the decision is to fail to reject H_0, indicating that at a 5 percent significance level, the constructed model fits the data used in this analysis.

5. Classification Accuracy Measures

Classification accuracy measures are presented as an additional complementary tool for assessing model performance. The results of the classification table are shown in Table 7 below.

Table 7. Results of the Classification Table

Predicted	**Observed**	
	$y = 1$	$y = 0$
$y = 1$	1,613	32
$y = 0$	558	113
Total	2,171	145

Based on the classification table, the accuracy is 74.53 percent, indicating that the model correctly classifies the employment status of educated women in 74.53 percent of all observations. Furthermore, a sensitivity of 74.3 percent indicates that the model correctly identifies 74.3 percent of employed educated women as being employed. Meanwhile, a specificity of 77.93 percent indicates that the model correctly classifies 77.93 percent of unemployed educated women as unemployed.

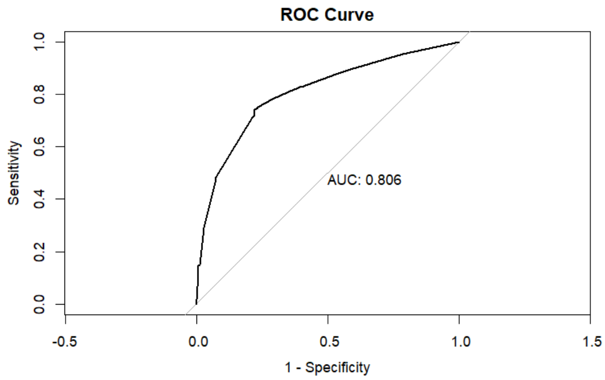

Figure 3. Area Under ROC Curve

Subsequently, the ROC curve results are presented in Figure 3. The area under ROC curve is 0.806, which falls within the range of 0.8 to 0.9. According to Hosmer et al. (2013), this value indicates that the model exhibits excellent discriminative ability in distinguishing between employed and unemployed educated women.

6. Model Interpretation

The tendency of the predictor variables on the employment status of educated women are presented through the estimated odds ratio $(\widehat{OR})$, which are shown in Table 5 and visualized in the odds ratio plot in Figure 4.

The odds ratio (OR) plot in the Figure 4 shows the effect of each independent variable on the probability of educated women being employed. The dashed vertical line at OR = 1 represents the absence of an effect. ICT use, age, marital status, and place of residence have odds ratios greater than one with 95% confidence intervals that do not cross one, indicating that these variables have a significant effect on the employment probability of educated women. In contrast, the training variable has a confidence interval that crosses one, suggesting that its effect is not significant.

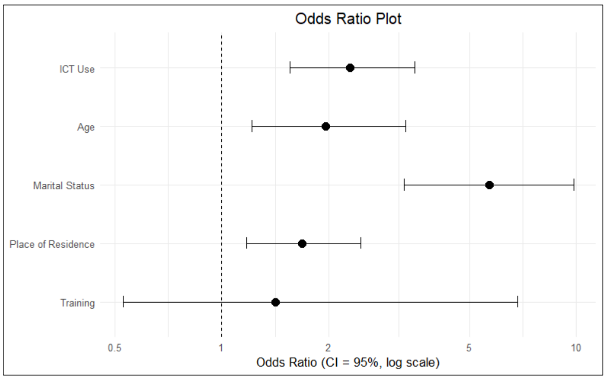

Figure 4. Odds Ratio Plot

As indicated by the pattern in the odds ratio plot, the following subsection presents a detailed interpretation of each predictor variable.

a) ICT Use

Based on the partial test results in Table 5, the variable of ICT use shows a test statistic value of $|W|$ = 4.134 > $Z_{(0,05/2)}$ = 1.96 with a p-value < 0.05, indicating that at a 5 percent significance level, ICT use has a significant effect on the employment status of educated women. The value of $\widehat{OR}$ for ICT use is 2.301, meaning that educated women who use three ICT devices are 2.301 times more likely to be employed compared to those who use fewer than three devices.

At present, almost all types of work are shifting toward technology-based systems (Potgieter, 2021). Aligns with Suhaib & Kartiasih (2024), who state that the use of mobile phones, the internet, and computers encourages women's participation in the labor market. Using all three devices simultaneously reflects stronger ICT skills. Proficiency in ICT enables women to more easily obtain information about job opportunities (Nikulin, 2017). Furthermore, ICT use facilitates participation in more flexible forms of work, such as digital based jobs or digital entrepreneurship. So that it can help women to balance career and household duties.

b) Age

The partial test for the age variable shows $|W|$ = 2.703 > $Z_{(0,05/2)}$ = 1.96 with a p-value < 0.05, indicating that at a 5 percent significance level, age significantly affects the employment status of educated women. The value of $\widehat{OR}$ is 1.964 meaning that educated women who are not in the young age group are 1.964 times more likely to be employed than those in the young age group (15–24 years).

According to Damayanti (2021), women of more mature ages, up to 64 years old, tend to have a higher likelihood of being employed. The younger age group represents a transitional period from education to the workforce (Gaffari & Handayani, 2019). Youth entering the labor market for the first time face greater challenges due to limited work experience and insufficient knowledge of available job opportunities. Furthermore, according to La Ode (2023), youth have high idealism in choosing jobs. In particular, those with higher educational attainment tend to be selective in order to obtain wages that are commensurate with their level of education.

c) Marital Status

The partial test for marital status shows $|W|$ = 6.277 > $Z_{(0,05/2)}$ = 1.96 with a p-value < 0.05, indicating that at a 5 percent significance level, marital status significantly affects the employment status of educated women. The value of $\widehat{OR}$ for this variable is 5.683, meaning that educated women who are ever married are 5.683 times more likely to be employed than those who unmarried.

This finding is consistent with Damayanti (2021) who found that women's tendency to work is higher among those who are ever married. This may occur because ever married women have higher financial needs than those who unmarried. In line with this, according to Andoh & Ampofo (2018) as cited in Shantika (2019), although women are often still positioned as having a greater role in domestic affairs than in economic activities, there has been a shift in perspective among educated women. They continue to choose to work even after marriage as a form of self-actualization and economic independence.

d) Place of Residence

The partial test for the residential area variable shows $|W|$ = 2.797 > $Z_{(0,05/2)}$=1.96 with a p-value < 0.05, indicating that place of residence significantly affects the employment status of educated women. The value of $\widehat{OR}$ is 1.688, meaning that educated women living in rural areas are 1.688 times more likely to be employed than those living in urban areas.

This result is consistent with Tsaniyah & Sugiharti (2021), who found that women in rural areas are more likely to work. Urban areas tend to have more intense job competition and higher migration inflows, which can reduce employment opportunities for women in cities.

e) Training

The training variable doesn't show a significant effect on the employment status of educated women. This result suggests that general participation in training is not sufficient to differentiate employment opportunities. Nevertheless, this training variable is retained in the model as an indicator of overall training participation, as training is conceptually regarded as an investment in skills and a component of human capital that is relevant in explaining employment decisions (McQuaid & Lindsay, 2005). The use of more detailed measures, such as specific types of training, may result in very small cell sizes given the relatively low proportion of respondents who participated in training (5,77 percent), thereby limiting the reliability of the inferred conclusions.

CONCLUSION

The application of Firth's logistic regression in this study shows that ICT use, age, marital status, and place of residence significantly influence the employment status of educated women in Riau Province in 2024. The tendency of being employed is higher among educated women who use three ICT devices, are not in the younger age group, ever married, and live in rural areas. Policymakers can design more effective initiatives to boost the employment participation of educated women through expanded ICT access and consideration of sociodemographic characteristics.

Firth's logistic regression was employed to obtain less biased parameter estimates under data conditions with potential imbalance between employed and unemployed categories. The model successfully identified key determinants of employment status among educated women, providing insights into both ICT and sociodemographic factors influencing women's labor participation.

This study has limitation. In this analysis, the training variable is operationalized as a general indicator of training participation. As a result, the analysis may not fully reflect differences across training characteristics. Future studies may explore these aspects in greater detail and also consider additional factors through primary data such as local or cultural norms that may influence employment decisions. (2022)

REFERENCE

Adriani, D., & Yustini, T. (2021). Anticipating the demographic bonus from the perspective of human capital in Indonesia. *International Journal of Research in Business and Social Science (2147-4478)*, *10*(6), 141–152. https://doi.org/10.20525/ijrbs.v10i6.1377

Amini, F., & Oktora, S. I. (2021). Comorbid of chronic kidney disease (CKD) patients who undergoing dialysis in Indonesia using firth logistic regression. *THE 2ND SCIENCE AND MATHEMATICS INTERNATIONAL CONFERENCE (SMIC 2020): Transforming Research and Education of Science and Mathematics in the Digital Age*, 020009. https://doi.org/10.1063/5.0041667

BPS. (2022). *Keadaan angkatan kerja Indonesia Agustus 2022*.

BPS. (2023). *Keadaan angkatan kerja Indonesia Agustus 2023*.

BPS. (2024a). *Indeks Pembangunan Teknologi Informasi dan Komunikasi 2024*.

BPS. (2024b). *Keadaan angkatan kerja Indonesia Agustus 2024*.

BPS. (2024c). *Keadaan angkatan kerja Riau Agustus 2024*.

BPS. (2024d). *Persentase penduduk usia 25 Tahun keatas dengan pendidikan SMA ke atas menurut jenis kelamin (Persen), 2024*. BPS. https://www.bps.go.id/id/statistics-table/2/MjE5OSMy/persentasependuduk-usia-25-tahun-keatas-dengan-pendidikan-sma-ke-atas-menurut-jeniskelamin--persen-

Damayanti, K. (2021). Determinan perempuan bekerja di Jawa Barat. *Jurnal Kependudukan Indonesia*, *16*(1), 55–66. https://doi.org/10.14203/jki.v16i1.428

Firth, D. (1993). Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, *80*(1), 27–38. https://doi.org/10.2307/2336755

Gaffari, A., & Handayani, D. (2019). Keputusan usia muda yang tidak bekerja dan tidak terikat pendidikan (nee) dan karakteristiknya di Indonesia. *Jurnal Ekonomi*, *22*(2), 76–91.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. Wiley. https://doi.org/10.1002/9781118548387

Koto, R., Ridwan, E., & Muharja, F. (2025). Analisis Partisipasi Angkatan Kerja Perempuan di Indonesia. *Jurnal Informatika Ekonomi Bisnis*, *7*(3), 607–612. https://doi.org/10.37034/infeb.v7i3.1221

Kurniasih, C. E., Tampubolon, D., & Ula, T. (2022). Analisis Pengaruh Indikator Pasar Tenaga Kerja Perempuan Terhadap Kemiskinan Antar Kabupaten/Kota di Provinsi Riau. *National Multidisciplinary Sciences*, *1*(4), 572–584. https://doi.org/10.32528/nms.v1i4.109

Kusumawardhani, N., Pramana, R., Saputri, N. S., & Suryadarma, D. (2023). Heterogeneous impact of internet availability on female labor market outcomes in an emerging economy: Evidence from Indonesia. *World Development*, *164*, 106182. https://doi.org/10.1016/j.worlddev.2022.106182

La Ode, M. H. (2023). Durasi Mencari Kerja Bagi Pekerja Usia Muda di Indonesia. *Jurnal Forum Analisis Statistik (FORMASI)*, *2*(2), 118–128. https://doi.org/10.57059/formasi.v2i2.38

Lusiyanti, L., & Wicaksono, P. (2020). The Impact of Education and Social Demographic Factors on Female Labor Force Participation in Indonesia. *Muwazah*, *12*(2), 219–236.

Mankiw, N. Gregory., & Harris, R. B. (1998). *Principles of microeconomics*. Dryden Press.

McQuaid, R. W., & Lindsay, C. (2005). The Concept of Employability. *Urban Studies*, *42*(2), 197–219. https://doi.org/10.1080/0042098042000316100

Mulugeta, G. (2021). The role and determinants of women labor force participation for household poverty reduction in Debre Birhan town, North Shewa zone, Ethiopia. *Cogent Economics & Finance*, *9*(1). https://doi.org/10.1080/23322039.2021.1892927

Ngoa, G. B. N., & Song, J. S. (2021). Female participation in African labor markets: The role of information and communication technologies. *Telecommunications Policy*, *45*(9), 102174.

https://doi.org/10.1016/j.telpol.2021.102174
Nikulin, D. (2017). The Impact of ICTs on Women's Economic Empowerment. In *Catalyzing Development through ICT Adoption* (pp. 15–24). Springer International Publishing. https://doi.org/10.1007/978-3-319-56523-1_2
Niu, L. (2020). A review of the application of logistic regression in educational research: common issues, implications, and suggestions. *Educational Review*, *72*(1), 41–67. https://doi.org/10.1080/00131911.2018.1483892
Potgieter, I. L. (2021). Surviving the Digital Era: The Link Between Positive Coping, Workplace Friendships and Career Adaptability. In *Agile Coping in the Digital Workplace* (pp. 57–78). Springer International Publishing. https://doi.org/10.1007/978-3-030-70228-1_4
Shah, G. H., Etheredge, G. D., Schwind, J. S., Maluantesa, L., Waterfield, K. C., Mulenga, A., Ikhile, O., Engetele, E., & Ayangunna, E. (2022). Firth's Logistic Regression of Interruption in Treatment before and after the Onset of COVID-19 among People Living with HIV on ART in Two Provinces of DRC. *Healthcare*, *10*(8), 1516. https://doi.org/10.3390/healthcare10081516
Shantika, F. (2019). *Analisis peran ganda pada perempuan terdidik dalam perspektif sosial - ekonomi dan etnis di Indonesia* [Thesis]. Universitas Brawijaya.
Suhaib, A. M., & Kartiasih, F. (2024). Pengaruh Teknologi Digital terhadap Partisipasi Perempuan dalam Angkatan Kerja di Provinsi Papua. *Jurnal Ketenagakerjaan*, *19*(2), 218–232. https://doi.org/10.47198/jnaker.v19i2.342
Tsaniyah, A. H., & Sugiharti, L. (2021). The determinants of women's work: a case study in East Java. *Jurnal Ilmu Ekonomi Terapan*, *6*(1), 66–81.
World Bank. (2011). *World Development Report: Gender Equality in Indonesia Improving*. World Bank. https://www.worldbank.org/en/news/press-release/2011/09/20/world-development-report-gender-equality-indonesia-improving
Yin, J., & Tian, L. (2014). Joint inference about sensitivity and specificity at the optimal cut-off point associated with Youden index. *Computational Statistics & Data Analysis*, *77*, 1–13. https://doi.org/10.1016/j.csda.2014.01.021
Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35.