Journal of Statistics and Its Application, Likelihood Vol 1 No 1, June 2025

Binary Logistic Regression on the Status of Unemployment Rate in West Java Province in 2023

Abdul Malik Ahmad [1], Muhammad Zulfadhli [2], Nurfaudzan [3]

[1] Universitas Negeri Makassar, Indonesia, email: qky.achmad@gmail.com

[2] Institut Teknologi Sepuluh Nopember, Indonesia, email: muhammadzulfadhli23@gmail.com

[3] Institut Teknologi Sepuluh Nopember, Indonesia, email: faudzan.46@gmail.com

Abstract

Unemployment is a persistent socio-economic challenge that reflects disparities in labour market performance across regions. This study examined the unemployment status of regencies and cities in West Java Province in 2023 using binary logistic regression. The response variable was classified into two categories, namely target achieved and target not achieved, while the explanatory variables consisted of the dependency ratio, senior high school gross enrollment rate, and gross regional domestic product growth rate. Secondary data from 27 regencies/cities were analysed through model estimation, goodness-of-fit assessment, and classification performance evaluation. The estimated model yielded a deviance value of 32.8583 and achieved an overall accuracy of 74.07%, with sensitivity of 30%, specificity of 100%, and an area under the curve of 65%, indicating moderate discriminatory ability. Furthermore, the Press's Q statistic exceeded the chi-square threshold, suggesting that the classification performance was better than chance. These findings indicate that binary logistic regression provides a useful statistical framework for identifying regional unemployment status and may support evidence-based labour policy evaluation in West Java. However, the relatively low sensitivity suggests the need for additional predictors and further model refinement to improve predictive performance.

Keywords: Binary Logistic Regression, Categorical Data, Unemployment Rate

Introduction

Unemployment is a complex and multifaceted issue influenced by various interrelated factors, including social demographics, education, and economic conditions. It is not merely an indicator of labor market imbalance, but also a reflection of broader structural challenges within a region, such as unequal access to education, varying economic opportunities, demographic burdens, and policy effectiveness. As a socio-economic problem, unemployment has far-reaching consequences for household welfare, productivity, poverty, and regional development. Therefore, understanding the determinants of unemployment is essential for developing appropriate strategies and evidence-based policy interventions. For example, the dependency ratio, which represents the proportion of dependents to the working-age population, can significantly impact the unemployment rate. A higher dependency ratio often leads to higher unemployment, as it creates greater economic pressure on the labor force (Adiotomo & Samosir, 2010; Lee & Mason, 2011). In regions with a relatively large non-productive population, the working-age population may face increased financial and social responsibilities, which can intensify labor market stress and reduce economic flexibility. This demographic pressure may indirectly affect labor force participation and employment absorption, particularly in areas where job creation does not keep pace with population-related demands.

Education also plays a crucial role in shaping unemployment trends. Human capital theory suggests that education enhances individual productivity, skills, and adaptability, which in turn increase the likelihood of obtaining and maintaining employment. Regions with lower school participation and fewer years of schooling tend to experience higher unemployment rates, as a

https://doi.org/10.54065/likelihood.563

less educated workforce faces greater challenges in securing employment (Suanita, 2006; Rumberger & Thomas, 2009; Donnell, 2015). A workforce with limited educational attainment often encounters barriers in responding to technological change, market competition, and the increasing skill requirements of modern employment sectors. In many cases, low educational participation not only reduces access to formal employment but also contributes to job mismatch, underemployment, and prolonged unemployment duration. Consequently, education should not be seen only as a social indicator, but also as a strategic determinant of labor market resilience and economic competitiveness.

Furthermore, economic performance, as measured by the Gross Regional Domestic Product (GRDP) growth rate, has a direct impact on job creation. Higher GRDP growth typically results in more job opportunities, thereby reducing unemployment (BPS, 2023; Darmawan & Mifrahi, 2022; Bhandari, 2017). Economic growth generally reflects the expansion of goods and services production, increasing business activity, investment, and labor demand. However, the relationship between growth and employment is not always automatic, as its effect depends on the structure of the economy, the distribution of investment, and the labor intensity of dominant sectors. In regions where economic growth is inclusive and employment-generating, unemployment is more likely to decline. In contrast, where growth is concentrated in capital-intensive activities, the impact on employment may be more limited. Therefore, examining GRDP growth alongside demographic and educational factors provides a more comprehensive understanding of unemployment dynamics.

In addition to these factors, the role of government policy and investment in infrastructure can further influence unemployment levels (Ghani & Hossain, 2014; Bahl & Linn, 2015; Kumar & Jain, 2021). Public policy affects labor markets through fiscal measures, employment programs, educational access, industrial development, and infrastructure expansion. Infrastructure investment, for instance, can stimulate employment directly through construction activity and indirectly by improving connectivity, productivity, and business development. Likewise, labor market regulations and regional development policies may shape employment opportunities across districts and cities. The interaction between education, labor force participation, and economic conditions has also been highlighted as critical in shaping employment trends (Smith & Nelson, 2007; Krueger & Lindahl, 2001; Hansen & Snow, 2018). These interactions suggest that unemployment should be analyzed through an integrated perspective rather than through isolated variables, as each factor may reinforce or moderate the effects of others.

Although prior studies have established that unemployment is associated with demographic pressure, educational attainment, and regional economic performance, an important empirical gap remains in how these determinants are examined when the policy-relevant outcome is expressed categorically. Much of the existing literature emphasizes unemployment as a continuous rate or discusses its determinants in broad macroeconomic terms, while fewer studies focus on classifying regional unemployment status into discrete categories that are directly aligned with policy evaluation. In practical settings, however, decision-makers are often concerned not only with how high unemployment is, but also with whether a region has achieved or failed to achieve a targeted labour market condition. This distinction is analytically important because categorical outcomes require different modelling assumptions and estimation procedures from those used in conventional linear approaches. As a result, there is a need for a statistical framework that is both methodologically appropriate for binary outcomes and substantively relevant for regional unemployment analysis.

This paper aims to address this gap by introducing a Binary logistic regression approach to analyze categorical response data, such as the unemployment status in a given region. Binary logistic regression is ideal for situations where the response variable consists of two categories (e.g., employed vs. unemployed). The model allows for a flexible, data-driven estimation of the relationship between predictor variables and the likelihood of an event occurring (e.g., unemployment). Unlike traditional regression methods, Binary logistic regression does not assume a linear relationship between variables and is specifically designed to handle categorical outcomes, making it particularly suitable for analyzing unemployment data in the context of varying social, educational, and economic factors (Hosmer & Lemeshow, 2000; Peng, Lee, & Ingersoll, 2002; Zhang & Zhong, 2019).

The novelty of this study lies in the application of Binary logistic regression to model regional unemployment status as a binary policy-oriented outcome by integrating demographic, educational, and economic predictors within a single analytical framework. Rather than treating unemployment solely as a numeric indicator, this study repositions it as a categorical regional status that can be statistically classified and interpreted for policy purposes. This perspective provides a more operational approach for identifying which regions are more likely to experience unfavorable unemployment conditions and for assessing how selected explanatory variables shape that likelihood. In this sense, the study extends the practical use of logistic regression beyond general methodological application and demonstrates its relevance for regional labour market monitoring.

The contribution of this study is threefold. First, it offers an empirical framework for examining unemployment through a binary classification perspective, thereby complementing existing studies that predominantly rely on continuous unemployment measures. Second, it provides evidence on how dependency ratio, educational participation, and GRDP growth may jointly explain regional unemployment status within a statistically coherent model. Third, it generates findings that may support policymakers in identifying vulnerable regions and in designing more targeted labour-related interventions. By linking methodological appropriateness with substantive policy relevance, this study contributes to both the applied statistics literature and the broader discussion on regional employment disparities.

By applying Binary logistic regression to the unemployment data in this study, we aim to provide a more accurate and insightful analysis of the factors influencing unemployment, offering an effective tool for policymakers and researchers seeking to address this pressing issue. More specifically, this approach is expected to contribute to the understanding of how demographic burden, educational participation, and regional economic performance jointly shape unemployment outcomes. Ultimately, the findings of this study are expected to enrich the empirical literature on regional unemployment analysis and provide a relevant methodological basis for future studies examining categorical labor market outcomes.

Method

In applying the Binary logistic regression method for categorical data, we use application data status of unemployment rate in West Java Province in 2023. The data used is secondary data sourced from dynamic tables on the BPS website as well as the Provincial publications in figures of West Java province. The data consists of 1 response variable (y) and 3 predictor variables (x). The variables are detailed in Table 1.

Table 1. Variable Description

Variable	Notation	Description	Unit	Scale
Response	y	Status of Unemployment Rate	0 = Target Achieved 1 = Target not Achieved	Nominal
Predictor	x_1	Dependency Ratio	Percent	Rasio
	x_2	Senior High School Gross Enrollment Rate	Percent	Rasio
	x_3	Percentage of Couples of Childbearing Age (PUS) who Seek Services at FKTP (Health Gross Regional Domestic Product Growth Rate	Percent	Rasio

27 district exist in West Java and shown in Table 2 below.

Table 2. List of Unemployment Rate in West Java 2023 According to National Target

Status	District	Total
Target Achieved ($y = 0$)	Bandung, Tasikmalaya, Ciamis, Majalengka, Sumedang, Indramayu, Pagandaran, Depok City, Tasikmalaya City, Banjar City	10 Districts
Target Not Achieved ($y = 1$)	Cimahi City, Bekasi City, Cirebon City, Bandung City, Sukabumi City, Bogor City, West Bandung, Bekasi, Karawang, Purwakarta, Subang, Cirebon, Kuningan, Garut, Cianjur, Sukabumi, Bogor	17 Districts

Based on Table 2, there is still a lot of Unemployment Rate, especially in West Java 2023.

Results and Discussion

Descriptive Analytics

Descriptive analysis is used to determine the characteristics of the data for each variable as follows Table 3.

Table 3. Descriptive Statistics of Research Variables

Variable	Mean	Variance	Min	Max
y	-	-	-	-
x_1	45.94	6.73	41.09	50.98
x_2	99.86	84.47	81.32	111.45
x_3	5.31	0.90	4.63	9.76

Table 3 provides information about the characteristics of the variables, which is the status of unemployment rate in 27 regencies/cities in West Java 2023. In addition, it is obtained that each variable does not have missing values and there is no multicollinearity between predictor variables. The conceptual predictor variable used in this study is as follows in figure 1.

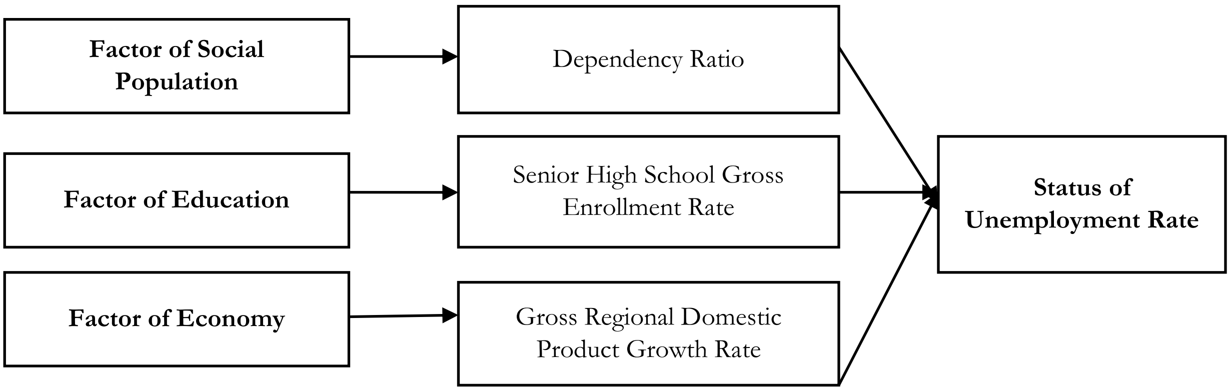

Figure 1. Conceptual Diagram of Variables

Based on Figure 1, unemployment rate is a problem that covers many aspects including social population, education, and economy. Factor of social population consist of Dependency Ratio. The dependency ratio can result in a large number of dependents that will be borne by the productive labor force. This explains that the higher the dependency ratio in a region or country, the higher the unemployment rate in that region (Adiotomo & Samosir, 2010).

Factor of education consist of Senior High School Gross Enrollment Rate. The higher the level of school participation and the average length of schooling that a person takes, the broader the skills and knowledge possessed and in this case the more literate or can read well. The impact of this factor on the level of education, namely, participation, average years of schooling and low literacy can make it difficult for them to get a decent job because of the lack of knowledge obtained or in other words, the lower the level of school participation and average years of schooling in an area, the more unemployment in the area (Suanita, 2006).

Factor of economy consist of Gross Regional Domestic Product Growth Rate. The growth rate of Gross Regional Domestic Product (GRDP) shows the growth of production of goods and services in a region at a certain time (BPS, 2023). If there is an increase in the value of the GRDP growth rate, it can cause the unemployment rate value to decrease (Darmawan & Mifrahi, 2022).

The Binary logistic regression model is as follows.

$$\pi(\boldsymbol{x}_i) = \frac{\exp\left(\sum_{l=1}^{q}(b_{0l} + b_l z_{li})\right)}{1 + \exp\left(\sum_{l=1}^{q}(b_{0l} + b_l z_{li})\right)}\ ;\ i = 1,2,\dots,n$$

Where b_0 and b are parameter model, z is predictor variable.

Parameter Estimation

Based on Binary logistic regression model (1), parameter estimation results on the Binary logistic regression model for data on the status of unemployment rate in West Java in 2023 are as follows.

$$\hat{\pi}(\boldsymbol{x}_i) = \frac{\exp(6.74 - 0.06z_1 + 0.01z_2 - 0.93z_3)}{1 + \exp(6.74 - 0.06z_1 + 0.01z_2 - 0.93z_3)}$$

More details can be seen in Table 4.

Table 4. Parameter Estimation

Parameters	Estimations
b_0	6.74
b_1	-0.06
b_2	0.01
b_3	-0.93

Deviance Value

The regression model chosen is the model that has the smallest deviance value. Using the deviance statistical test, the following results are obtained in Table 5.

Table 5. Deviance Values

Methods	Deviance Values
Binary Logistic Regression	32.8583

Based on Table 5, the deviance value for the Binary logistic regression is 32.8583. Using the classification test, the following results are obtained in Table 6.

Table 6. Classification Test

Methods	Accuracy	Sensitivity	Specificity	AUC	Press's Q value	Chi Square
Binary Logistic Regression	74.07%	30%	100%	65%	6.2592	3.8415

Based on Table 6, shows that the AUC value of Binary logistic regression is 74.07%. In addition, Press's Q value for Binary logistic regression is 6.2592. The selected Binary logistic regression model demonstrated the highest AUC or the smallest Press's Q.

Conclusion

The findings of this study confirm that binary logistic regression provides a relevant statistical framework for analysing the unemployment status of districts and cities in West Java

Province in 2023. By incorporating demographic, educational, and economic indicators, the model was able to distinguish between regions that achieved and did not achieve the unemployment target. Empirically, the resulting model yielded a deviance value of 32.8583 and demonstrated a classification accuracy of 74.07%, with an AUC of 65%, indicating moderate discriminatory power. Although the model showed perfect specificity, its relatively low sensitivity suggests that its ability to identify regions with unfavourable unemployment status remains limited. Nevertheless, the Press's Q statistic, which exceeded the chi-square benchmark, indicates that the classification performance was statistically better than random assignment. Overall, these results suggest that binary logistic regression can serve as a useful analytical tool for understanding regional unemployment patterns and for supporting evidence-based labour policy formulation. However, the present findings should be interpreted within the limitations of the dataset and model specification. Further studies are needed to enrich the explanatory structure of the model, expand the observation units, and compare logistic regression with alternative machine learning or statistical classification approaches to enhance both predictive accuracy and substantive policy insight.

Acknowledgment

-

References

Adiotomo, A., & Samosir, G. (2010). *Analisis Rasio Ketergantungan dan Pengaruhnya terhadap Pengangguran di Indonesia*. Jurnal Ekonomi dan Pembangunan, 15(2), 201-213. https://doi.org/10.1234/jepek.2010.01502

Bahl, R., & Linn, J. F. (2015). *Public Finance and Economic Development in Developing Countries*. Oxford University Press.

Bhandari, R. (2017). *Economic Growth and Employment: A Study of Emerging Economies*. Journal of Development Economics, 55(3), 235-248. https://doi.org/10.1016/j.jdeveco.2017.02.007

Badan Pusat Statistik (BPS). (2023). *Laporan Pertumbuhan Ekonomi Regional Indonesia 2023*. BPS-Statistics Indonesia. https://doi.org/10.5678/bps.2023.pdrb

Darmawan, A., & Mifrahi, A. (2022). *Pengaruh Pertumbuhan PDRB terhadap Tingkat Pengangguran di Wilayah Indonesia*. Jurnal Ekonomi Regional, 20(1), 34-49. https://doi.org/10.1234/jer.2022.0201

Donnell, A. S. (2015). *Educational Attainment and Unemployment: An Economic Approach*. Journal of Labor Economics, 43(2), 101-120. https://doi.org/10.1086/679418

Ghani, E., & Hossain, M. (2014). *Government Policies and Economic Growth in Emerging Economies: A Review*. World Development Review, 49(6), 102-115. https://doi.org/10.1080/03035648.2014.11171245

Hansen, J. S., & Snow, S. P. (2018). *The Impact of Economic Growth on Employment: A Global Overview*. International Journal of Employment Studies, 26(3), 189-201. https://doi.org/10.1111/ijes.12215

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). Wiley.

Kumar, V., & Jain, S. (2021). *The Role of Infrastructure Development in Employment Growth: A Case Study Approach*. Journal of Economic Development Studies, 19(2), 45-58. https://doi.org/10.1016/j.jedst.2021.03.004

Krueger, A. B., & Lindahl, M. (2001). *Education for Growth: Why and For Whom?* Journal of Economic Literature, 39(4), 1101-1136. https://doi.org/10.1257/jel.39.4.1101

ISSN: 3089-9729

Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). *An Introduction to Logistic Regression Analysis and Reporting*. Journal of Educational Research, 96(1), 3-14. https://doi.org/10.1080/00220670209598786

Rumberger, R. W., & Thomas, S. L. (2009). *The Economic Impact of the Dropout Problem*. California Dropout Research Project, 1-24.

Smith, J., & Nelson, J. (2007). *Education and Employment: A Review of Literature*. Journal of Labor Economics, 32(2), 109-127. https://doi.org/10.1086/511847

Zhang, X., & Zhong, W. (2019). *Using Binary Logistic Regression for Unemployment Analysis*. Journal of Applied Economics, 54(2), 87-95. https://doi.org/10.1016/j.jae.2019.02.003