Al-Jabar: Jurnal Pendidikan Matematika
Volume 15, Number 02, 2024, Pages 593 – 610
DOI: 10.24042/ajpm.v15i2.24518

Utilize imagery and crowdsourced data on spatial employment modelling

Novi Hidayat Pusponegoro[1]*, Ro'fah Nur Rachmawati[2], Maria A. Hasiholan Siallagan[1], Ditto Satrio Wicaksono[1]
[1] Politeknik Statistika STIS, Indonesia
[2] Universitas Pertahanan Indonesia, Indonesia
✉ novie@stis.ac.id *

Article Information
Submitted Oct 02, 2024
Accepted Dec 16, 2024
Published Dec 30, 2024

Keywords
Adaptive Bi-square;
Big Data;
Geographically Weighted Regression (GWR);
Labor Absorption.

Abstract

Background: Spatial employment modeling investigates employment distribution, patterns, influencing factors, neighboring area impact, and regional policy efficacy. Conventional studies often rely on traditional data sources, which may overlook critical employment-related phenomena. In 2022, Java recorded the lowest labor absorption rate in Indonesia, necessitating a new approach.
Aim: This study combines imagery, crowdsourced data, and official statistics to identify factors influencing labor absorption in Java Island.
Method: Geographically Weighted Regression (GWR) was employed to account for spatial effects in the data.
Results: The model reveals that nighttime light intensity in urban and agricultural areas, along with environmental quality, significantly enhances labor absorption across Java. Internet facilities, universities, and the number of micro and small industries also positively influence most districts/cities.
Conclusion: Incorporating new data sources offers valuable insights for understanding employment patterns and can enrich employment research frameworks.

INTRODUCTION

The eight goals of the Sustainable Development Goals (SDGs) are designed to promote sustained, inclusive, and sustainable economic growth, full and productive employment, and decent work for all (United Nations, 2017). Therefore, every country should endeavor to achieve positive economic growth and ensure that financial progress results in the creation of decent and fulfilling jobs while not compromising the environment. Labor absorption represents a pivotal regional development indicator. An imbalance between labor force growth and labor absorption will inevitably result in elevated unemployment rates (Todaro, 1998, 2006).

Indonesia is a country with the fourth largest population in the world, exhibiting a high rate of population growth. A large population results in an imbalance between supply and demand in the labor market, which may lead to a reduction in labor absorption when the number of available jobs is limited. It is challenging to achieve maximum labor absorption in Indonesia's free market economy (Mankiw, 2006). In a recent report, Statistics Indonesia indicated that 8.425 million individuals, representing 5.86 percent of the labor force, had not yet secured employment by 2022. This percentage is notably elevated in comparison to other countries in Southeast Asia (ASEAN, 2023).

How to cite	Pusponegoro, N. H., Rachmawati, R. N., Siallagan, M. A. H., & Wicaksono, D. S. (2024). Utilize imagery and crowdsourced data on spatial employment modelling. *Al-Jabar: Pendidikan Matematika*, *15*(2), 593-610.
E-ISSN	2540-7562
Published by	Mathematics Education Department, UIN Raden Intan Lampung

According to Statistics Indonesia, Java Island attracted the largest number of migrants compared to other islands in 2022, yet demonstrated the least effective absorption capacity. Conversely, Java Island represents the epicenter of the Indonesian economy, industry, and economic activity. It is crucial to identify the factors influencing labor absorption in Java. The multifaceted issue of labor absorption in Java Island necessitates an investigation into the factors that can enhance its efficacy. Furthermore, the issue of labor absorption necessitates adjustments to accommodate the advent of the Society 5.0 era (Kolade & Owoseni, 2022; Tavares et al., 2022). The advent of the Fourth Industrial Revolution has brought with it the necessity for humans to coexist with technology. The impact of this phenomenon may vary depending on the extent to which each country is able to utilize it effectively. The positive impacts include increased productivity and the emergence of new job roles (Grybauskas et al., 2022). One adverse consequence is the displacement of human labor by technology, which is perceived as more cost-effective (Filippi et al., 2023; Hötte et al., 2023).

The study of labor absorption determinants is a well-established field of research, given its pivotal role in the development of labor markets. Factors influencing labor absorption in Java include improvements in micro and small industrial units, industrial output, and input costs (Wulandani & Winarti, 2024). Additionally, labor issues are influenced by technological (Wang et al., 2024; You et al., 2024) and environmental aspects (Kjellström et al., 2019; Totouom et al., 2024). From a socio-economic perspective, Java Island displays a notable degree of spatial heterogeneity, with each region exhibiting distinctive characteristics (Rizki Oktarina et al., 2023; Wicaksono et al., 2024). Similarly, the rate of labor absorption may vary across regions due to differences in economic conditions, industrial composition, or demographic factors. Spatial modeling in labor absorption patterns and trends that vary geographically can be analyzed with the Geographically Weighted Regression (GWR) (Siallagan & Pusponegoro, 2024). The objective of spatial modeling in employment research is to analyze the patterns and trends of employment that vary in a geographical context. The consideration of spatial effects on labor absorption allows for the implementation of targeted and effective enhancement strategies, as policies enacted in a given region can also impact neighboring regions. It can therefore be argued that an analysis of labor absorption incorporating spatial elements will provide a more comprehensive interpretation.

Previous studies on labor absorption have employed official statistics generated by national statistical organizations through censuses, surveys, or other conventional data sources (Putri & Sudarsono, 2019; Jannah et al., 2021; Alisyahbana & Anwar, 2022). In the current era of big data, the size of databases and the technology used to process them have both developed rapidly. This has led to the availability of high-dimensional variables in modeling approaches. In conjunction with the advancement of technology, novel data sources have emerged that can be harnessed for analytical purposes. This data is described by the term 3V, namely volume, velocity, and variety (Favaretto et al., 2020). In other words, the data sources are numerous, generated and processed rapidly, and contain a substantial quantity of information. The data is more cost-effective, requires less time to collect, offers greater precision in results, and can be used to generate a variety of information (Jamarani et al., 2024). Accordingly, this study employs a novel data source, namely imagery and crowdsourced data, in the context of spatial labor absorption modeling.

The aim of this research is to combine imagery and crowdsourced data with official statistical data in order to determine the factors that affect employment in Java. Consequently, this research can serve as a basis for the government to develop policies aimed at enhancing employment and as a point of reference for future research that employs crowdsourced and imagery data as a data source. The following is a description of the structure of this paper: Section 2 provides an overview of the methodology employed, outlining the data sources, data preprocessing techniques, and analytical tools utilized in the investigation. Section 3 describes and discusses the findings of the research project. Section 4 presents the conclusions drawn from the research.

METHODS

Instruments

This study employs secondary data from 119 districts/cities in six provinces on the island of Java. The data employed in this study is cross-sectional in nature, obtained from the BPS in 2022, and comprises both remote sensing imagery data and other forms of data. Moreover, this research employs crowdsourced data, which is defined as big data collected by a large number of individuals (Siallagan & Wijayanto, 2023). The dependent variable in this study is labor absorption. Labor absorption can be defined as the number of jobs that have been filled, as observed from the number of people who are gainfully employed (Weir-Smith & Dlamini, 2024).

Table 1. Descriptions of Variables Obtained from Official Statistics

No	Variables	Operational Definition	Source
1.	Labor Absorption	Number of people 15 years and over who are employed	SAKERNAS 2022
2.	Number of Universities	Number of private and public universities	PODES 2021
3.	Number of Micro and Small Industries	Total number of micro and small industries (with less than 20 workers)	PODES 2021

Table 2. Descriptions of Imagery and Crowdsourced Data Variables Used

No	Variables	Description	Source	Resolution Spatial	Period Update	Unit	Band
1.	Night Time Light (NTL)	Detection of economic activity distribution	NOAA-VIIRS (Payne Institute, 2021)	463.83 m	Monthly	nano Watt/sr/cm^2	avg_rad
2.	Network Performance	Download and upload speed of mobile devices with mobile	GEE Community Catalog	610.8 m	Annual	kbps	avg_d_kbps avg_u_kbps

		connection type (e.g. WiFi, Ethernet)	of Ookla Apps (gee-community, 2023)				
3.	Normalized Difference Vegetation Index (NDVI)	Detection area vegetation	Sentinel-2 (SUHET, 2013)	10 m	5 days	Index	Band 8 (NIR) Band 4 (Red)
4.	Classification Cover Land	Land classification urban and land agriculture	MODIS (Friedl & Sulla-Menashe, 2022)	500 m	Annual		LC_Type1

The independent variables of this study comprise nighttime light (NTL) as a proxy for economic activity, which is subsequently classified according to land cover types into urban NTL (X_1) and agricultural NTL (X_2). Furthermore, network performance as an approach to technology (X_3), Normalized Difference Vegetation Index (NDVI) (X_4) which is an indicator of environmental greening, the number of universities (X_5) as an approach to infrastructure educarion, the number of micro and small industries (X_6) that represent business factors. A detailed description of each research variable can be found in Table 1 and Table 2 above.

Preprocessing Data

Remotely sensed satellite imagery data is prepared through GEE. Preprocessing on the Sentinel-2 satellite includes cloud selection and cloud masking to produce a cloud-free image set. Cloud masking detects clouds in each region, removes the clouds, and replaces them with the median pixel value of the study time frame. The bands of the cloud-free image were then composited to obtain the NDVI index. The NDVI index is calculated as a ratio between the red (R) and near-infrared (NIR):

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{1}$$

In the next stage, data reduction is performed to obtain statistical values that represent all satellite images in 1 year (Putri et al., 2022). The NTL variable uses summation reduction to show the total night light intensity of an area. The NDVI variable uses median reduction because it is more resistant to outlier values due to cloud cover on the Sentinel-2 satellite in some areas. While the network performance variable uses average reduction to show the average speed of a region.

Furthermore, the MODIS satellite was used during the calculation of the NTL variable to obtain the NTL Urban and NTL Agriculture classifications. The MODIS satellite can classify land cover into 17 classes. This study uses the classification results from the Annual International Geosphere-Biosphere Program (IGBP) classification (LC_Type1). Furthermore, classifications with a value of 13 were used as urban classifications, and values of 10,12,14 were used as agricultural classifications (Table 3). These classifications were used to see the difference in the effect of night light in urban areas on labor absorption compared to night light in agricultural areas.

Table 3. Descriptions of the MCD12Q1.061 LC_Type1 Classification Used

Value	Description
10	Grassland: dominated by herbaceous annuals (<2 m)
12	Agricultural land: at least 60% of the area is agricultural land
13	Urban and built-up land: at least 30% impervious surface area, such as materials
14	Farmland/Natural Vegetation Mosaic: 40-60% small-scale-cultivation mosaic with vegetation

Source: SUHET (2013)

Data Analysis

Geographically Weighted Regression (GWR) is a technique used to model spatially varying relationships in data, allowing for a more nuanced understanding of the relationships between variables change across different geographic locations. GWR is a spatial model that can capture spatial variations locally (Lu et al., 2014) can be written as:

$$Y_i = \sum_{j}^{p} \beta_{ij}(u_i, v_i)X_{ij} + \varepsilon_i, \tag{2}$$

of the i-th observation and parameter $\beta_{ij}(u_i, v_i)$ is the function of (u_i, v_i) the i-th observation.

GWR extends traditional regression analysis by incorporating spatial variability into the model. While traditional regression assumes that relationships between variables are uniform across space, GWR allows these relationships to vary by location, providing a more detailed picture of spatial dynamics. Unlike traditional regression, where parameters are estimated globally, GWR estimates parameters locally, allowing them to vary spatially. GWR involves fitting a series of local regression models, one for each geographic location. The goal is to capture the spatial variation in the relationships between the dependent variable and independent variable, thus GWR estimates the parameter using Weighted Least Square (WLS) as follows:

$$\boldsymbol{\hat{\beta}(u_i, v_i) = [X^T W(u_i, v_i) X]^{-1} X^T W(u_i, v_i) Y,} \tag{3}$$

where $W(u_i, v_i)$ is a matrix of weights specific to the i-th observation with (u_i, v_i) is the coordinate of its geographical location. The observations nearer to the i-th observation are given greater weight than observations further away.

Parameter estimation in the GWR model requires a weighting matrix, where weights are given to the data based on their proximity to the observation location. There are common weight matrices such as Gaussian and Bi-square Kernel. In the instance, Adaptive Bi-square is a weight that is influenced by different bandwidths for each observation location with the following formula:

$$\boldsymbol{w_{ij} = \left\{ \left[1 - \left(\frac{d_{ij}}{b_{i(q)}} \right)^2 \right]^2 , \; d_{ij} < 0,} \tag{4}$$

where $\boldsymbol{d_{ij}}$ is the distance between the i-th and the j-th observation which is represented with the Euclidean distance. And, $\boldsymbol{b_{i(q)}}$ is the bandwidth value that will limit the number of nearest neighbors affecting the kernel. The greater the bandwidth value of a predictor variable, the more global the influence of that variable on the response variable.

The bandwidth may be either defined by a given distance, or a fixed number of nearest neighbors from the analysis location. The optimal number of nearest neighbors of the kernel function is determined by the smallest *Akaike Information Criterion Corrected* (AICc). The formula of AICc is as follows:

$$\boldsymbol{AIC_c} = \boldsymbol{2n(\hat{\sigma}) + n(2\pi\,) + n\left\{\frac{n + tr(S)}{n - 2 - tr(S)}\right\}}, \tag{5}$$

where $\boldsymbol{n}$ is the total observations, $\hat{\sigma}$ is the standard deviation of the error, $\boldsymbol{tr(S)}$ denotes the trace of the hat matrix, L is the model likelihood, and k is the number of parameters. The smaller the AICc value produced by the kernel function, the better the weight (Fotheringham et al., 2002).

Golden Search is a technique to find the optimum bandwidth that minimizes the AIC along the maximum and minimum interval of observation locations. Furthermore, for the models to be compared, the following criteria need to be tested (Mohammadinia et al., 2017):

1. R-squared as a measure of goodness of fit.
2. The Simultaneous Test has a p-value <0.05, meaning that there are independent variables that significantly affect the response variable.
3. Variance Inflation Factor (VIF) < 7.5 indicates there is no multicollinearity between independent variables.
4. The Kolmogorov statistic value is insignificant to indicate normally distributed residuals.
5. The results of Breusch Pagan tests are significant at the $\alpha = 5\%$.

Thus, this research utilizes the Adaptive Bi-square weights with bandwidth selection using Golden Search. Generally, this research process begins with the collection of variables sourced from official statistics as well as imagery data and crowdsource data. Subsequently, for the dependent variables, their distribution is examined through thematic maps and Moran's I testing to determine the presence of spatial effects.

Meanwhile, for variables sourced from imagery and crowdsource data, statistical calculations are conducted to generate the values of each variable, followed by visualization to understand the characteristics of each area. These variables are aggregated to obtain representative values for each district/city. Next, modeling is carried out starting from the formation of an OLS (Ordinary Least Squares) model. Once the OLS model is formed, assumptions are tested, including multicollinearity detection, normality, spatial autocorrelation, and spatial heterogeneity. If all assumptions are met, the OLS model can be used. However, if there is a violation of the spatial heterogeneity assumption, modeling with OLS will result in biased estimates.

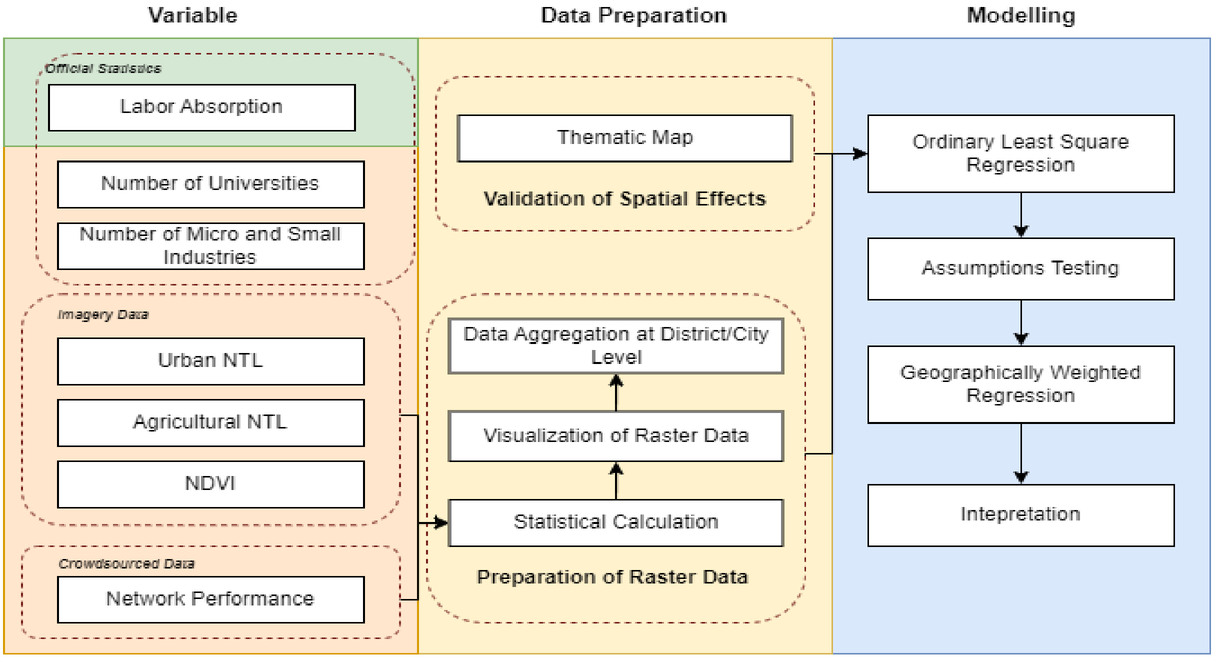


Figure 1. Research Stages

Spatial heterogeneity testing aims to determine variance heterogeneity among observations especially spatial diversity for each observed region (Anselin, 1988). This test uses the *Breusch-Pagan* statistic with the following hypothesis:

$H_0 = \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2,$

$H_1 = There\ is\ at\ least\ one\ \sigma_i^2 = \sigma^2.$

The test statistics can be written as follows:

$$\boldsymbol{BP} = \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{f}^{\boldsymbol{T}}\boldsymbol{Z}(\boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{Z})^{-\mathbf{1}}\boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{f} \sim \boldsymbol{\chi}^{\mathbf{2}}, \quad (6)$$

where $f = \left(\frac{e_i^2}{\sigma^2} - 1\right)$ with e_i is the *least square residuals* for the *i*-th observation, then $Z_{n\ x\ (p+1)}$ is the independent variable matrix containing the normalized vector $\boldsymbol{Z}$ for each observation. If $BP > \chi_p^2$the null hypothesis is rejected, identifying that there is spatial heterogeneity, then the GWR model can be applied. Therefore, modeling continues using a GWR model to accommodate spatial heterogeneity. The final stage is the interpretation of the GWR model. The research stages can be seen in Figure 1.

RESULTS AND DISCUSSION

Result

This study employs descriptive and inferential analysis using R, QGIS, and MGWR 2.2 software. Descriptive analysis is applied to identify spatial patterns in labor absorption and to provide an overview of imagery and crowdsourced data used as independent variables. The distribution of labor absorption across Java Island in 2022 is displayed in Figure **2**. Among districts and cities, Kepulauan Seribu has the lowest labor absorption, while Bogor Regency has the highest. Additionally, districts/cities in West Java with high labor absorption tend to be surrounded by other districts/cities with similarly high levels of labor absorption. This pattern indicates the presence of positive spatial autocorrelation in data. Beyond spatial autocorrelation, the labor absorption variable also exhibits spatial heterogeneity, as reflected in the variation in values among neighboring districts/cities.

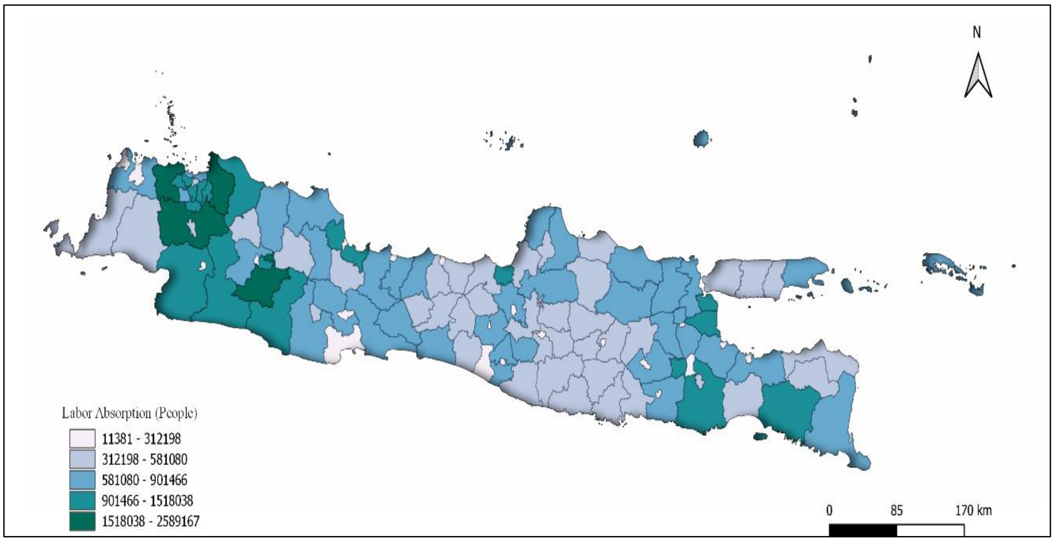


Figure 2. Distribution of Labor Absorption by District/City in 2022

Before modeling to determine the factors influencing labor absorption, independent variables strongly correlated with labor absorption are needed. These variables can be obtained from big data and crowdsourced data. The imagery data used in this study are from remote sensing using MODIS, VIIRS-NOAA, and Sentinel-2. The study also utilizes crowdsourced data *i.e.* average network download and upload speeds from *Ookla* applications.

The imagery and crowdsourced data in this study are provided in raster format. Raster data, which consist of pixels with individual values, may contain multiple bands in the same pixel, offering comprehensive information about a region's characteristics. Visualizations of raster data extracted for West Java Province are shown in Figure 3, Figure 4, and Figure 5. In these figures, the "labor absorption" variable represents aggregated labor absorption at the district/city level. For analytical consistency, variables derived from raster data are also presented at the district/city level.

Figure 3 shows the distribution of NDVI (Normalized Difference Vegetation Index) in West Java. NDVI reflects an area's vegetation health and density, with high values indicating robust, dense vegetation, often used as a proxy for environmental greening (Kim & Kim, 2020). Since most Indonesian farmers are based on Java Island, higher NDVI values likely support increased agricultural productivity and labor absorption. For instance, Sukabumi, Subang, and Tasikmalaya Regency have high NDVI values and corresponding high labor absorption.

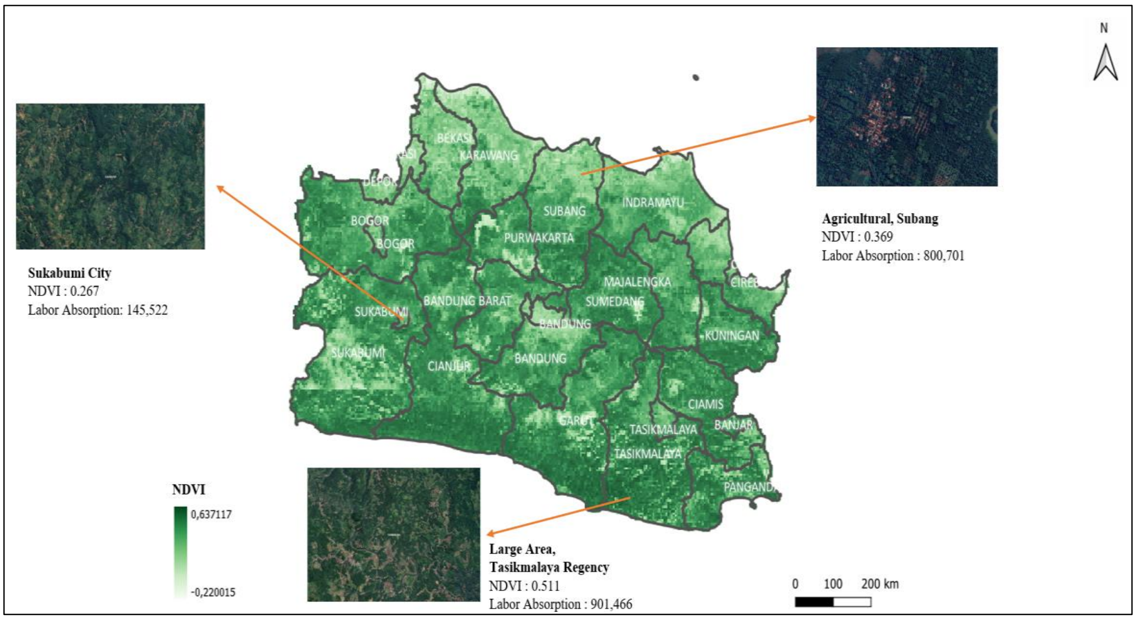


Figure 3. Distribution of NDVI

Next, Figure 4 shows the upload and download speed tests on mobile devices conducted by crowd in West Java. It can be seen that the distribution of internet speeds in West Java is remains uneven. Some big cities like Bekasi, Depok, Bandung, and Bogor have higher average network performance compared to other areas. Additionally, the increased network performance is positively associated with labor absorption in these cities.

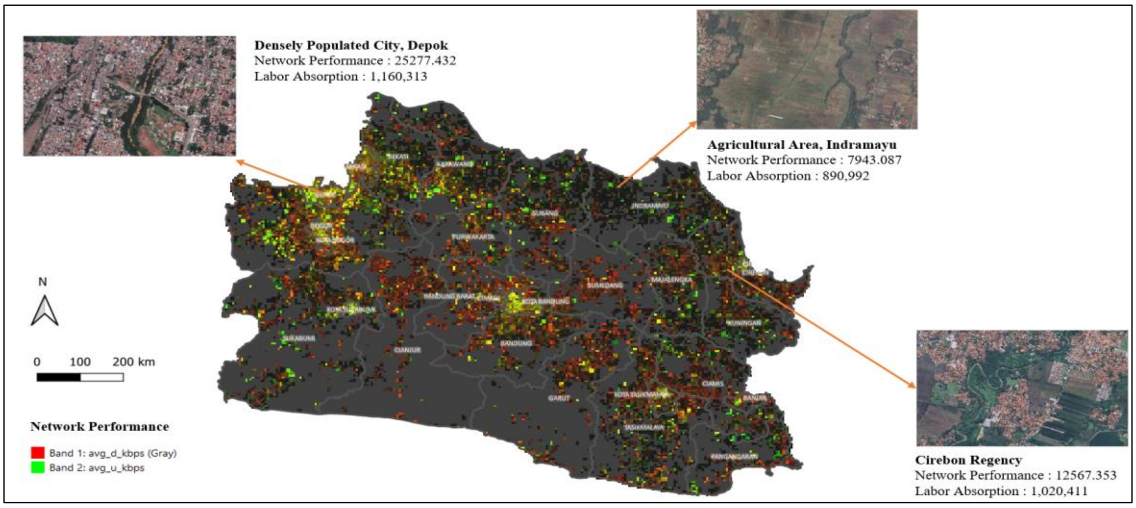


Figure 4. Distribution of Network Performance

Figure 5 illustrates the distribution of NOAA-VIIRS Night-Time Light (NTL) values across Java Island, classified into urban and agricultural areas based on MODIS satellite data. NTL values are often used as a proxy to assess economic progress in a region. As shown in Figure 5, most areas in West Java are classified as agricultural land rather than urban land. However, higher NTL values in both urban and agricultural areas are associated with higher labor absorption levels. This relationship is also evident from the significant and positive Pearson correlation coefficients of both variables with labor absorption. A notable finding from this classification is that the influence of urban NTL on labor absorption is greater than that of agricultural NTL. For example, the city of Bekasi, predominantly urban, has higher labor absorption compared to Subang, which is mainly agricultural. This is also reflected in the Pearson correlation value between Urban NTL and labor absorption, which is 0.6874, higher than the Pearson correlation for Rural NTL, which is only 0.4261.

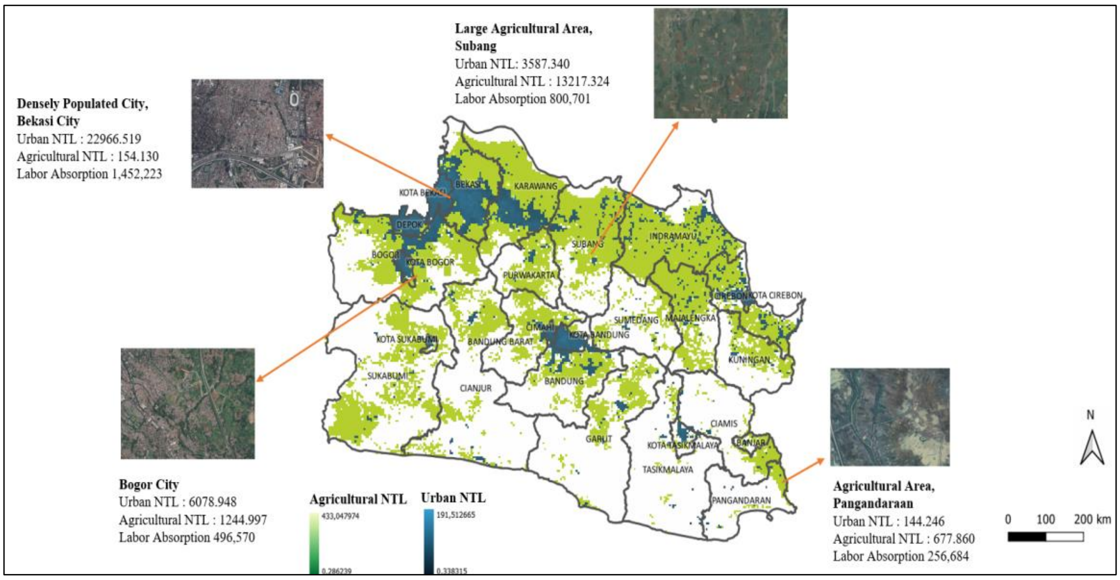


Figure 5. Distribution of NTL in Agricultural and Urban Areas

The variables from big data and crowdsourced data will then be integrated with official statistical data for modeling purposes. To ensure consistency for modeling, all variables must be standardized to the same spatial level. Imagery-based independent variables, such as the NDVI, are available at a 10m x 10m resolution, while labor absorption data from official statistics is only available at the district/city level. Consequently, all raster data are aggregated to the district/city level.

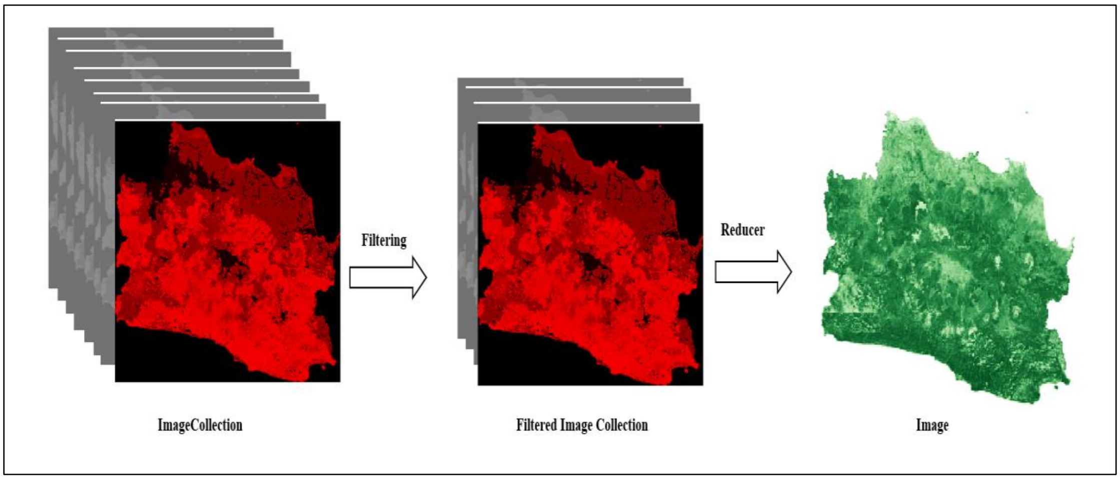
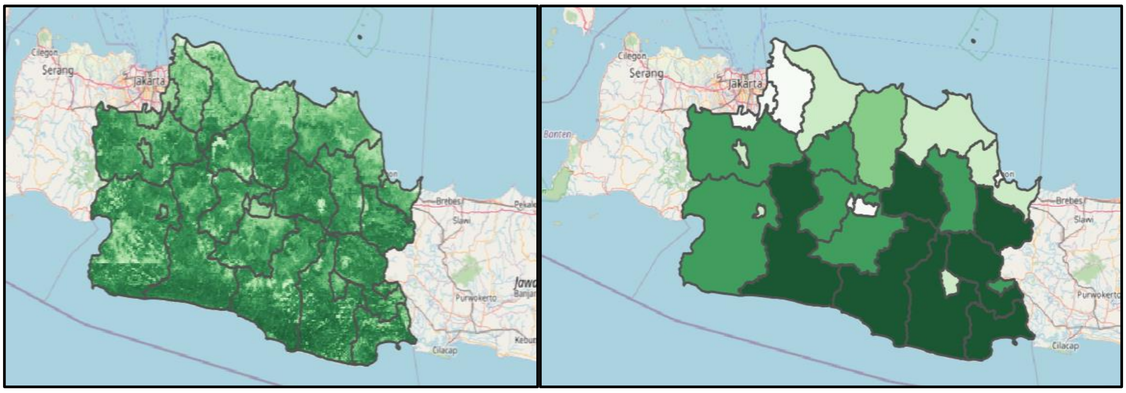


Figure 6. Raster Data Aggregation Process

Figure 6 and Figure 7 display the raster data aggregation process. Raster data from Google Earth Engine (GEE) is presented as an Image Collection, containing a series of satellite images over time. This collection is filtered to cover the study period, January 1, 2022, to December 31, 2022. Using GEE's Reducer function, statistical summaries for the entire Image Collection during the study period yield a single Image.



(a) (b)

Figure 7. NDVI Index a) before and b) after aggregation

Using GEE's Reduce Region function, zonal statistics are then applied, aggregating each pixel based on administrative boundaries, here at the district/city level. An example of zonal statistics application is shown in Figure 7. After zonal statistics, each district/city is represented by a single NDVI value. This process is repeated for all remote sensing-based independent variables, preparing them for modeling. Aggregated results are presented in Table 4.

Table 4. The Aggregated Imagery Data Variables at the District/City Level

No	District/city	NDVI	Network Performance	Agricultural NTL	Urban NTL
1.	Pangandaran	0.494089	11935.52	677.8609	144.2467
2.	Banjar	0.416137	16070.88	681.9071	213.8073
3.[a]	Tasikmalaya	0.511772	11834.77	1309.162	379.4618

Displayed 3 out of 119 district/city data

Due to scale differences across imagery data-derived variables, all variables are standardized. Standardization is performed when variable variances differ greatly, enabling equal weighting and facilitating interpretation and comparison. The assumption of non-multicollinearity is also tested, with each independent variable having a VIF value below 7.5, meeting the non-multicollinearity (see Table 5).

Table 5. Assumption of Non-multicollinearity

Notation	Variable	VIF
X_1	Urban NTL	2.340
X_2	Agricultural NTL	1.449
X_3	Network Performance	4.095
X_4	NDVI	3.124
X_5	Number of Universities	2.379
X_6	Number of Micro and Small Industries	1.548

Tests for spatial effects were then conducted. Sequential spatial heterogeneity tests using the Breusch-Pagan test show p-values below the 5% significance level, indicating spatial heterogeneity in labor absorption across Java Island (see Table 6). This heterogeneity could introduce bias in OLS estimation; therefore, the GWR model is applied to account for these effects. The model assumes a normally distributed dependent variable, which is confirmed by nonsignificant normality test statistics (see Table 6).

Table 6. Assumption Testing

Assumption	Test	p-value
Spatial Heteroskedasticity	*Breusch-Pagan*	0.0000
Normality	*Kolmogorov Smirnov*	0.3293

Due to the violation of heteroskedasticity in the OLS model, particularly spatial heteroskedasticity, the GWR model can be applied and evaluated. Table 7 compares the performance of different models in terms of their AICc and R^2 values, under two different conditions: "Without Imagery and Crowdsourced Data" and "With Imagery and Crowdsourced Data." It can be seen that incorporating imagery and crowdsourced data leads to a significant improvement in both the OLS and GWR models, as indicated by lower AICc values and higher R^2 values. This suggests that imagery and crowdsourced data provide valuable information that enhances the predictive power of the models. The model performance also emphasizes that the GWR model provides a lower AICc value and a higher R-squared value compared to the OLS model.

Table 7. Model Performance

Scenario	**Model**	**Evaluation**	
		AICc	$\mathbf{R^2}$
Without Imagery and Crowdsourced Data	OLS	293.313	0.347
	GWR	274.343	0.568
With Imagery and Crowdsourced Data	OLS	171.267	0.786
	GWR	155.121	0.841

In addition, the GWR model consistently outperforms the OLS model in both scenarios, highlighting the importance of addressing for spatial heterogeneity in the data. This demonstrates that the GWR model is both more accurate and more efficient than the OLS model. Based on the AICc and R^2 values, the GWR model with imagery and crowdsourced data is the best-performing model, offering the most accurate and reliable predictions in this analysis, and is thus selected as the best model. A summary of the estimation results of the GWR model formed can be found in Table 8.

Table 8. Summary Statistics of GWR Parameter Estimation Results

Notation	Variable	Mean	STD	Min	Median	Max
X_0	Intercept	-0.022	0.121	-0.176	-0.053	0.129
X_1	Urban NTL	0.692	0.030	0.633	0.690	0.740
X_2	Agricultural NTL	0.383	0.081	0.251	0.382	0.500
X_3	Network Performance	0.042	0.157	-0.167	0.011	0.226
X_4	Normalized Difference Vegetation Index	0.330	0.091	0.231	0.288	0.450
X_5	Number of Universities	0.230	0.081	0.129	0.184	0.337
X_6	Number of Micro and Small Industries	0.156	0.068	0.086	0.125	0.393

Discussion

In Table 8, it can be observed that the Urban NTL and Agricultural NTL variables have positive estimated coefficients, and at the 5% significance level, they significantly impact all districts/cities in Java Island. This means that an increase in nighttime light intensity in urban and agricultural areas significantly enhances labor absorption. A higher nighttime light intensity indicates increased economic activities in an area (Rybnikova, 2022). This economic growth subsequently demands more labor, thus boosting labor absorption in that region. Additionally, it is noticeable that the range of regression coefficients for Urban NTL (0.633 – 0.740) is greater than that of Agricultural NTL (0.251 – 0.500). This indicates that urban economic activities have a greater impact on increasing labor absorption compared to agricultural economic activities. Therefore, to enhance labor absorption, the government can boost economic activities in a district/city starting from urban areas first. This is also supported by the presence of spatial autocorrelation and spatial dependency, indicating that a policy will affect surrounding areas. This means that increasing economic activities starting from urban areas will also impact agricultural areas.

Furthermore, the network performance variable has regression coefficients ranging from -0.167 to 0.266. This variable has a negative coefficient in 56 districts/cities in Java Island, but its value is not significant at the 5% significance level, so its effect can be disregarded. On the other hand, this variable has a positive and significant impact on 42 districts/cities in Java Island. This means that in those districts/cities, an increase in network performance will increase labor absorption. Network performance can depict a region's accessibility to information technology facilities. The easier and faster a region is connected to the internet, the more rapid the technological development in that area (UNCTAD, 2019). Technological advancements bring changes to market dynamics where workers' skills will improve, thus increasing opportunities for entering the labor and creating new job opportunities (Hötte et al., 2023). This also aligns with the Solow economic theory stating that technology adoption will lead to labor efficiency and drive output. Increased output will affect increasing labor absorption.

The NDVI index variable has positive estimated coefficients ranging from 0.231 to 0.450, and it is significant in all districts/cities in Java Island. This means that the higher the NDVI value of an area, the greater the labor absorption available in that area. The NDVI value is often used as a measure of agricultural productivity. This index also has a tangible positive relationship, thus affecting agricultural productivity positively (Nafarin & Novitasari, 2023). Higher productivity requires more inputs, in this case, labor, to meet higher production levels. Therefore, labor absorption increases.

Next, the number of universities variable has estimated coefficients ranging from 0.129 to 0.337 and significantly impacts 53 districts/cities in Java Island. This means that in those districts/cities, an increase in the number of universities will enhance labor absorption. An increase in the number of universities will result in job creation, thereby enhancing labour absorption. Moreover, universities aim to produce quality educated manpower (Lauder & Mayhew, 2020). Thus, besides increasing the quantity of labor, increasing universities will also improve the quality of the labor in an area.

The Number of Micro and Small Industries variable has estimated coefficients ranging from 0.086 to 0.393 and is significant in 81 districts/cities in Java Island. This means that in those districts/cities, an increase in the number of small and micro-industries can absorb a larger labor. This is in line with research Amalia & Woyanti (2020), which indicate that larger business units will create new job opportunities for the population. Small and micro-industries are labor-intensive, meaning they require many workers in their production processes. Furthermore, they are also more flexible than large industries, thus facilitating the labor absorption process (Koeswahyono et al., 2022).

Figure 8 depicts the significant independent variables that influence labor absorption across districts and cities in Java Island. These factors are essential to consider, as they have statistically proven to increase labor absorption. The variables that significantly and globally impact all districts in Java Island are Urban NTL (X_1) and Agricultural NTL (X_2). Therefore, to enhance labor absorption, the government can stimulate economic activities, initially in urban areas and subsequently expanding to agricultural zones.

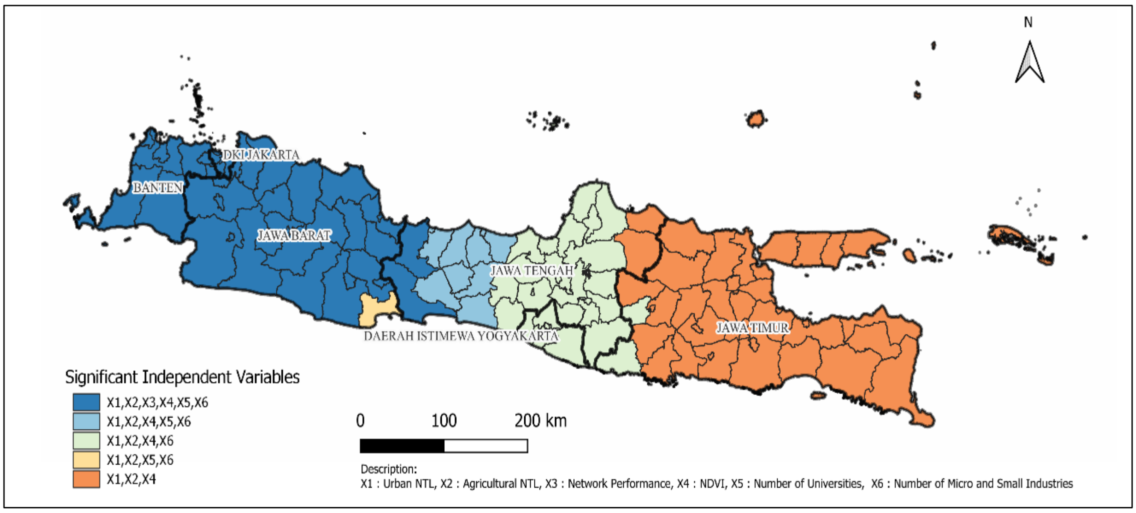


Figure 8. The Significant Independent Variables Influencing Labor Absorption by District/City

In addition to identifying the variables that globally impact all districts in Java, the GWR model also indicates variables that are significant only in specific districts. This feature is an advantage of the GWR model over OLS regression. Previous studies using OLS often considered Java as a homogeneous economic unit, applying a generalized approach to factors affecting labor absorption. For instance, Wulandani & Winarti, (2024) identified the Number of Micro and Small Industries as a primary driver across all provinces in Java. In contrast, our findings reveal that in almost all districts of East Java Province, this variable does not significantly impact labor absorption. This highlights the necessity of area-specific policies rather than a one-size-fits-all approach, as each area has unique characteristics.

In urban regions such as the provinces of Banten, DKI Jakarta, and West Java, labor absorption is significantly influenced by network performance (X_3). This implies that in these areas, the development and equitable distribution of technological and informational infrastructure, including internet access, must be improved to boost labor absorption. However, in other areas, network performance is not significant, suggesting that local governments in those regions could focus on other independent variables. For example, increasing the number of micro and small industries (X_6) in districts across Central Java could enhance labor absorption. Additionally, improving educational facilities (X_5), particularly higher education, could raise both the quantity and quality of labor in certain districts of Central Java (highlighted in light blue on the map) and in one district in West Java (Pangandaran, in yellow color on the map). Lastly, efforts to improve environmental quality through NDVI (X_4) across Java Island are crucial to support higher productivity and absorption in the agricultural sector, with the exception of Pangandaran.

Implication

This research successfully combines the use of imagery data and crowdsourced data with official statistics issued by the government. The use of imagery data and crowdsourced data can provide timely, dynamic, easily accessible, and cost-effective data. Imagery data can also provide proxy indicators such as NTL that strongly correlate with a country's economy, provide latent indicators such as NDVI as an environmental indicator, and provide data contributed by many people, such as internet speed data. Therefore, the use of imagery data can be said to offer significant opportunities to streamline and enrich the data analysis process.

CONCLUSIONS

This research demonstrates that the combination of imagery and crowdsourced data with official statistics can enhance the analysis of labor absorption determinants in Java Island in 2022. The Geographically Weighted Regression model indicates that the amount of nighttime light intensity in urban and agricultural areas, as well as environmental quality, have a significant positive impact on labor absorption in all districts/cities on Java Island. Additionally, an increase in the number of universities, micro and small industries, and network performance has been found to significantly enhance labor absorption in the majority of districts and cities on Java Island. These relationships entail spatial autocorrelation and spatial heterogeneity effects, which imply that government policies implemented in one area will be more efficient and targeted, as they will impact neighboring areas.

ACKNOWLEDGMENT

We would like to thank the reviewers and editors for their suggestions.

AUTHOR CONTRIBUTIONS STATEMENT

The author contributions to this research are as follows: Novi Hidayat Pusponegoro developed the research idea and designed the methodology, including the application of the Geographically Weighted Regression (GWR) model. They also processed data from official sources, imagery, and crowdsourced datasets and authored the introduction and conclusion sections of the article. Ro'fah Nur Rachmawati validated the data and performed statistical analyses using R, QGIS, and MGWR software. They contributed to drafting and editing the discussion section, connecting the findings to government policies on labor absorption. Maria A. Hasiholan Siallagan managed the collection of crowdsourced and imagery data and performed data preprocessing. They contributed to writing the methodology section, specifically focusing on raster data integration and aggregation processes. Ditto Satrio Wicaksono supported the research by providing literature and references on spatial modeling and big data. They were responsible for visualizing the data, including creating maps of distribution and spatial analysis graphics.

REFERENCES

Alisyahbana, A. N. Q. A., Fatmawati, & Anwar, A. I. (2022). Determinant Analysis of Labor Absorption in the Manufacturing Industry Sector in Sulawesi Island. *Advances in Economics, Business and Management Research, 205*. https://doi.org/https://doi.org/10.2991/aebmr.k.220107.041

Amalia, D., & Woyanti, N. (2020). The Effect of Business Unit, Production, Private Investment, and Minimum Wage on the Labor Absorption in the Large and Medium Industry 6 Provinces in Java Island. *Media Ekonomi Dan Manajemen*, *35*(2), 206. https://doi.org/10.24856/mem.v35i2.1550

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic. https://doi.org/10.1007/978-94-015-7799-1

ASEAN. (2023). *Asean Statistical Highlights 2023*.

Favaretto, M., de Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLoS ONE*, *15*(2). https://doi.org/10.1371/journal.pone.0228987

Filippi, E., Bannò, M., & Trento, S. (2023). Automation technologies and their impact on employment: A review, synthesis and future research agenda. *Technological Forecasting and Social Change, 191*. https://doi.org/10.1016/j.techfore.2023.122448

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: Analysis of Spatially Varying Relationship*. Wiley.

Friedl, M., & Sulla-Menashe, D. (2022). MODIS/Terra+ Aqua land cover type yearly L3 global 500m SIN grid V061. *NASA EOSDIS Land Processes DAAC*, *10*. https://doi.org/10.5067/MODIS/MCD12Q1.061

gee-community. (2023). *Speedtest® by Ookla® Global Fixed and Mobile Network Performance Maps. Based on analysis by Ookla of Speedtest Intelligence® data for [DATA TIME PERIOD]. Provided by Ookla and accessed [DAY MONTH YEAR]. Ookla trademarks used under license and reprinted with permission.* https://gee-community-catalog.org/projects/speedtest/

Grybauskas, A., Stefanini, A., & Ghobakhloo, M. (2022). Social sustainability in the age of digitalization: A systematic literature Review on the social implications of industry 4.0. *Technology in Society*, *70*. https://doi.org/10.1016/j.techsoc.2022.101997

Hötte, K., Somers, M., & Theodorakopoulos, A. (2023). Technology and jobs: A systematic literature review. *Technological Forecasting and Social Change*, *194*. https://doi.org/10.1016/j.techfore.2023.122750

Jamarani, A., Haddadi, S., Sarvizadeh, R., Haghi Kashani, M., Akbari, M., & Moradi, S. (2024). Big data and predictive analytics: A sytematic review of applications. *Artificial Intelligence Review*, *57*(7). https://doi.org/10.1007/s10462-024-10811-5

Jannah, E. M., Rachmadhan, A. A., Meidaliyantisyah, M., & Hendra, J. (2021). The Labor Force Absorption Improvement of Micro and Small Enterprises (MSEs) in Indonesia. *Jurnal Ekonomi Dan Studi Pembangunan*, *13*(1), 72. https://doi.org/10.17977/um002v13i12021p072

Kim, Y. J., & Kim, E. J. (2020). Neighborhood greenery as a predictor of outdoor crimes between low and high-income neighborhoods. *International Journal of Environmental Research and Public Health*, *17*(5). https://doi.org/10.3390/ijerph17051470

Kjellström, Tord., Maître, Nicolas., Saget, Catherine., Otto, Matthias., & Karimova, Takhmina. (2019). *Working on a warmer planet : the effect of heat stress on productivity and decent work*. International Labour Organization.

Koeswahyono, I., Maharani, D. P., & Liemanto, A. (2022). Legal breakthrough of the Indonesian job creation law for ease, protection, and empowerment of MSMEs during the COVID-19 pandemic. *Cogent Social Sciences*, *8*(1). https://doi.org/10.1080/23311886.2022.2084895

Kolade, O., & Owoseni, A. (2022). Employment 5.0: The work of the future and the future of work. *Technology in Society*, *71*. https://doi.org/10.1016/j.techsoc.2022.102086

Lauder, H., & Mayhew, K. (2020). Higher education and the labour market: an introduction. *Oxford Review of Education*, *46*(1), 1–9. https://doi.org/10.1080/03054985.2019.1699714

Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S. (2014). Geographically weighted regression with a non-Euclidean distance metric: A case study using hedonic house price data. *International Journal of Geographical Information Science*, *28*(4), 660–681. https://doi.org/10.1080/13658816.2013.865739

Mankiw, N. G. (2006). *Principles of Economics* (4th ed.). Cengage Learning.

Mohammadinia, A., Alimohammadi, A., & Saeidian, B. (2017). Efficiency of geographically weighted regression in modeling human leptospirosis based on environmental factors in Gilan province, Iran. *Geosciences (Switzerland)*, *7*(4). https://doi.org/10.3390/geosciences7040136

Nafarin, N. A., & Novitasari, N. (2023). Relationship between Normalized Difference Vegetation Index (NDVI) and Rice Growth Phases in Danda Jaya Swamp Irrigation Area Regency Barito Kuala. *IOP Conference Series: Earth and Environmental Science*, *1184*(1). https://doi.org/10.1088/1755-1315/1184/1/012019

Payne Institute. (2021). *VIIRS Nighttime Light*. https://eogdata.mines.edu/products/vnl/#monthly

Putri, N., & Sudarsono, H. (2019). The Analysis of Factors that Affect Labor Absorption in Natural Rubber Plantation. *Ekuilibrium: Jurnal Ilmiah Bidang Ilmu Ekonomi*, *14*(2).https://doi.org/https://doi.org/10.24269/ekuilibrium.v14i2.2019.pp120-135

Putri, S. R., Wijayanto, A. W., & Sakti, A. D. (2022). Developing Relative Spatial Poverty Index Using Integrated Remote Sensing and Geospatial Big Data Approach: A Case Study of East Java, Indonesia. *ISPRS International Journal of Geo-Information*, *11*(5). https://doi.org/10.3390/ijgi11050275

Rizki Oktarina, C., Aprianto, E., & Hidayati, N. (2023). Modeling the Open Unemployment Rate of Regency/City in West Java Province in 2021 using Spatial Autoregresive Moving Average and Spatial Durbin Model. *Journal of Statistics and Data Science*, *2*(2). https://ejournal.unib.ac.id/index.php/jsds/index

Rybnikova, N. (2022). *Nighttime Lights as a Proxy for Economic Performance of Regions*. https://doi.org/10.3390/rs14040825

Siallagan, M. A. H., & Pusponegoro, N. H. (2024). Spatial Regression Approach to Modelling Poverty in Java Island 2022. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *18*(3), 1765–1778. https://doi.org/10.30598/barekengvol18iss3pp1765-1778

Siallagan, M. A. H., & Wijayanto, A. W. (2023). Sentiment Analysis and Topic Modelling on Crowdsourced Data. *Indonesian Journal of Artificial Intelligence and Data Mining*, *7*(1). https://doi.org/10.24014/ijaidm.v7i1.24777

SUHET. (2013). *Sentinel-2 User Handbook*.

Tavares, M. C., Azevedo, G., & Marques, R. P. (2022). The Challenges and Opportunities of Era 5.0 for a More Humanistic and Sustainable Society—A Literature Review. *Societies*, *12*(6). https://doi.org/10.3390/soc12060149

Todaro, M. P. (1998). *Economic Development in The Third World*.

Todaro, M. P. (2006). *Economic Development* (9th ed.).

Totouom, A., Nkengfack, H., & Ngameni, J. P. (2024). Effect of dependence on natural resources on employment quality: Insights from African countries. *Australian Journal of Agricultural and Resource Economics*, *68*(4). https://doi.org/https://doi.org/10.1111/1467-8489.12587

UNCTAD. (2019). *The Impact of Rapid Technological Change on Sustainable Development*.

United Nations. (2017). *Work of the Statistical Commission pertaining to the 2030 Agenda for Sustainable Development*.

Wang, J., Tian, Z., & Sun, Y. (2024). Digital Economy, Employment Structure and Labor Share. *Sustainability*, *16*. https://doi.org/https://doi.org/10.3390/su16219584

Weir-Smith, G., & Dlamini, S. (2024). Hit the Road: Spatial Characteristics of Labor Absorption in South Africa. *Professional Geographer*, *76*(3), 331–342. https://doi.org/10.1080/00330124.2023.2300801

Wicaksono, D. S., Nuriyah, S., Fajritia, R., Yuniarti, N. P. N., Priatmadani, P., Amelia, L., & Berliana, S. M. (2024). Modeling Factors Affecting Educated Unemployment on Java Island using Geographically Weighted Poisson Regression Model. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *18*(1), 0615–0626. https://doi.org/10.30598/barekengvol18iss1pp0615-0626

Wulandani, F. I., & Winarti, A. S. (2024). Factors Influencing Labor Absorption in Micro and Small Industries on Java Island (2017-2022): An Analysis of Provincial Minimum Wage, Industrial Units, and Economic Output. *South Asian Journal of Social Studies and Economics*, *21*(9), 21–28. https://doi.org/10.9734/sajsse/2024/v21i9878

You, J., Xu, X., Liao, D., & Lin, C. (2024). International comparison of the impact of digital transformation on employment. *Journal of Asian Economics*, *95*. https://doi.org/10.1016/j.asieco.2024.101820