DOI: https://doi.org/10.34123/jurnalasks.v17i2.856

Jurnal Aplikasi Statistika & Komputasi Statistik
Vol. 17 No. 2, December 2025
e- ISSN: 2615-1367, p-ISSN:2086-4132

Available online at https://jurnal.stis.ac.id/

Estimating Economic Activity Using Geospatial Big Data in East Java, Indonesia: Relative Spatial GDP Index Approach

Rifqi Ramadhan[1*], I Made Satria Ambara Putra[2] , Taufiq Agung Kurniawan[3], Fitri Kartiasih[4], Raden Muaz Munim[5], Somethea Buoy[6]

[1]BPS-Statistics Indonesia, Jakarta, Indonesia, [2]BPS-Statistics Kotawaringin Barat Regency, Pangkalan Bun, Indonesia, [3]BPS-Statistics Dharmasraya Regency, Pulau Punjung, Indonesia, [4]Politeknik Statistika STIS, Jakarta, Indonesia, [5]Department of Computer Science, Hamad Bin Khalifa University, Qatar, [6]Department of the Statistical Information Services, National Institute of Statistics, Ministry of Planning, Cambodia
**Corresponding Author: E-mail address: rifqi.ramadhan@bps.go.id*

ARTICLE INFO

Article history:
Received 17 October, 2025
Revised 21 November, 2025
Accepted 28 November, 2025
Published 31 December, 2025

Keywords:
East Java; Geospatial Big Data; GRDP; Relative Spatial GDP Index; Remote Sensing; Robust Regression

Abstract

Introduction/Main Objectives: GRDP serves as a fundamental indicator for assessing regional economic performance in Indonesia and plays a critical role in development planning. **Background Problems:** Conventional GRDP measurement in Indonesia relies on survey-based approaches, which are time-consuming, costly, and provide limited spatial detail. **Novelty:** This study introduces a Relative Spatial GDP Index (RSGI) constructed from geospatial big data such as remote sensing and point of interest (POI) to estimate GRDP more granular in East Java. This approach represents the first geospatial data driven GRDP index developed at such fine spatial resolution in Indonesia. **Research Methods:** Four weighting schemes were applied to generate RSGI variations, which were then evaluated through regression modeling against official GRDP. They are equal weight, pearson correlation, spearman correlation, and principal component analysis (PCA). **Finding/Results:** The RSGI PCA produced the best performance (RMSE = 0.73047; MAE = 0.48185; MAPE = 7.00%; R^2 = 0.7618). PCA weight outperformed other weight by capturing shared variance and generating objective weights that better represent spatial economic intensity. The RSGI PCA demonstrates a strong and significant correlation with GRDP at the sub-district level and provides a robust tool for fine-scale economic estimation.

1. Introduction

Economic growth is one of the international development agendas included in the eighth point of the 2030 Sustainable Development Goals (SDGs). This agenda was also adopted by the Indonesian government in the National Medium-Term Development Plan (RPJMN) for the 2020–2024 period [1]. The expected economic growth target for the 2020–2024 period will increase by an average of 5.7–6.0 percent per year [1]. National economic growth, according to data from the Central Bureau of Statistics (BPS) of the Republic of Indonesia as of February 2023, grew by 5.31% [2]. This achievement is still below the economic growth target to be achieved in the prepared RPJMN. From this description, the Gross Domestic Product (GDP) has a special place, which is the focus of the SDGs development. According to [2], GDP is an important indicator to determine economic conditions in a country over a certain period.

Economic development is a series of efforts to reduce unemployment, improve people's welfare, and minimize inequality in people's incomes. In measuring the progress or health of a country's economy, the Gross Domestic Product (GDP) is usually used as an important indicator both of current prices and based on constant prices (fixed prices). Meanwhile, to measure the economic progress or health of a province, district, or city, the Gross Regional Domestic Product (GRDP) is used as one of the indicators. GRDP is one indicator that can measure economic activity in a region. GRDP is prepared using three approaches, namely production, income, and expenditure, and is presented based on current prices and constant (real) prices. Although the components of presenting GRDP data are different for each calculation approach, conceptually, these three approaches will produce the same figure [3]. GRDP at current prices (nominal GRDP) is prepared based on prices prevailing during the calculation period and aims to look at the structure of the economy. Meanwhile, GRDP at constant (real) prices is prepared based on prices in the base year [2]. GRDP at current prices shows the ability of economic resources produced by a region. GRDP at constant prices can be used to show the rate of economic growth as a whole or for each category from year to year. GRDP per capita at current prices shows the value of GRDP per population, while GRDP per capita at constant prices is used to determine the real economic growth per capita population of a region [3].

GRDP using the expenditure approach is calculated based on all components of final demand, such as Final Household Consumption Expenditure, Final Consumption Expenditures for Non-Profit Institutions serving Households, Government Final Consumption Expenditure, Gross Domestic Fixed Capital Formation, Changes in Inventories, and Net Exports. In the final household consumption expenditure (PKRT) component, there are various data sources used to compile GRDP data, such as the National Socio-Economic Survey (SUSENAS). Within one year, SUSENAS data collection was carried out twice, namely in March and in September [4]. However, to carry out periodic monitoring in a shorter time and in more detail in terms of area or granularity, SUSENAS data cannot accommodate this, considering that SUSENAS is conducted every six months and the level of presentation is only at districts or cities.

On the other hand, the calculation of GRDP uses a production approach using various data sources collected by BPS as well as compilation results from various other data sources. For example, in calculating GRDP for the agricultural, forestry, and fisheries components, one of the data sources used comes from the agricultural census, which is carried out once every ten years. Conventional survey data collection conducted in Indonesia for the calculation of GRDP has various weaknesses, such as SUSENAS, which is conducted every six months with a method that is limited in scope, costs a lot, and takes a long time. If monitoring in the short term, periodically, and in more detail in terms of SUSENAS data areas cannot provide it, a supporting data source with a faster and more efficient method is needed. In addition, similar problems can also be found in other GRDP-forming data sources, such as the SKLNP, which is carried out every year on a quarterly basis, as well as the agricultural census, which is carried out once every 10 years.

Today, many studies have turned to alternatives data sources such as satellite imagery, which is the result of remote sensing, and other big data sources, such as Point of Interest (POI) from OpenStreetMap (OSM). The use of other big data sources has been carried out by [5], and it can be concluded that POI can effectively differentiate industrial and commercial areas and therefore has the potential to improve GRDP mapping. Previous research has shown that Visible Infrared Imaging Radiometer Suite (VIIRS) remote sensing data that captures Nighttime Light Intensity (NTL) has a positive correlation to GDP in a region [6]–[10]. In addition, Sentinel-2 Multispectral Instrument Level-2A, which captures the Normalized Difference Vegetation Index (NDVI), influences GDP and has a significant correlation [11]–[16]. Other composite indices, such as the Normalized Difference Water Index (NDWI), detect a reduction in water sources associated with GDP and population growth [17]. Then, the Normalized Difference Built-Up Index (NDBI) relates to the area of built-up land and the expansion of urban areas [18]. Sentinel-5P satellite imagery is designed to detect pollution emissions in an area such as CO, which is closely related to urban areas and economic growth [15], [19], and NO2 shows results that have a strong correlation with the value of GDP [20]. and SO2 emissions, along with NO2, are indicated to have a relationship to the intensity of energy use and economic growth [21]. Therefore, pollution emissions from CO, SO2, and NO2 can be used to determine the GDP value of an area because they have an association with economic growth. Images derived from MODIS can detect the magnitude of the ground surface temperature in an area during the day and night where the ground surface temperature will have a significant impact due to an increase in population, industrial activity, infrastructure development, and socio-economic activities [22]. The accurate use of elevation data, namely the Digital Elevation Model (DEM) derived from ASTER satellite imagery, can provide a scientific and socio-economic point of view, and provide important things in geomorphology [23]. Another data source

originating from OpenStreetMap (OSM) can be utilized as a POI data source, which is a geospatial data source with geographic locations and descriptions based on categories and can effectively differentiate industrial and commercial areas, so that it has the potential to improve the precise GDP mapping of secondary and tertiary sectors [5]. Thus, the source of big data from satellite imagery and POI can be used as an alternative data source in data collection to estimate the GDP value of a region using relatively low costs, a short time frame, and continuous updating as needed.

Furthermore, recent literature shows a growing shift toward leveraging diverse geospatial big data sources to strengthen the ability of models to map economic indicators at the sub-national scale. The previously dominant linear regression approaches are now considered less adequate due to their limitations in capturing non-linear relationships and inter-variable spatial interactions. This limitation has encouraged researchers to adopt more flexible Machine Learning (ML) methods, including Random Forest which have demonstrated notable improvements in GDP estimation, particularly when combining multisource remote sensing data with POI data for regional economic activity mapping in China [24]. Empirical evidence shows that the Random Forest model demonstrated strong predictive performance, achieving R^2 values of 0.95 for non-agricultural GDP. Alongside these developments, Deep Learning (DL) approaches, especially Convolutional Neural Networks (CNN) applied to high-resolution daytime satellite imagery and nightlights data have shown promise in identifying subtle spatial patterns linked to economic activity in China. Using an attention-enhanced VGG-16 architecture to extract image features, this approach produced a strong predictive result in predicting localized economic activity, reaching an R^2 value of 0.71 [25]. Incorporating remote sensing data has also been shown to substantially improve the ability for GRDP mapping, particularly within dense urban environments such as NTL from NPP-VIIRS dan built-up area extraction from Sentinel-2 images, where the predicting of GRDP shows the R^2 values to 0.82 in Shandong Province, China [26]. In addition, downscaling mapping techniques are increasingly adopted to redistribute official aggregate statistics into finer spatial units. Recent advances that combine NTL, POI, and multispectral features within multi-scale deep learning frameworks have further demonstrated the potential to generate GDP maps that are more stable, detailed, and representative of real economic conditions [27]. However, despite the rapid growth of geospatial and machine learning based approaches, most existing studies remain centered on improving prediction accuracy rather than constructing a spatially explicit economic index that can capture relative economic intensity at finer administrative levels, which is essential for evidence-based local policy and development planning. Therefore, this study develops a spatial GDP index that is crucial for capturing economic activity as the alternatives data to provides a more robust evidence base for development-oriented policymaking.

For the record, the alternatives data can only be used as support for SUSENAS data and cannot be used as the essence or core of data from the field obtained from SUSENAS. Provinces in Java dominate the performance and structure of the Indonesian economy spatially, with a contribution of 56.48% and a growth of 5.31% in 2022 [2]. East Java Province was chosen as the case study in this research because East Java Province is in second place after DKI Jakarta in terms of achieving the GRDP level in Java Island (BPS, 2022). To anticipate the weakness of GRDP data collection, which still uses conventional surveys, the contribution of this research is to implement the utilization of integrating remote sensing data and POI data as geospatial big data sources a combination that has not yet been explored in the Indonesian context to estimate GRDP at a smaller granular level down to the sub-district level and map it so that it can be easily interpreted. This granularity data is crucial because economic activities and development patterns often exhibit significant heterogeneity within a single district or city. Relying solely on aggregate district/city level data masks critical localized disparities in wealth, infrastructure needs, and the effectiveness of local policies, thereby limiting the government's ability to conduct targeted intervention, policy formulation, and balanced regional planning. Then, this research also offers the first empirical contribution of its kind in Indonesia, demonstrating the potential of geospatial big data to generate spatially detailed economic indicators beyond the limitations of conventional survey-based GRDP statistics.

2. Material and Methods

2.1. Study Area

This study was conducted in East Java Province, Indonesia, with 2022 as the reference year. East Java consists of 38 regencies and municipalities and 666 sub-districts, covering an area of approximately 48,000 km², making it the largest province on Java Island. During the period 2020–2022, East Java recorded the second-highest Gross Regional Domestic Product (GRDP) at constant 2010 prices, after

DKI Jakarta, which motivated its selection as the study area. Moreover, the province exhibits substantial economic diversity across manufacturing, agro-industry, agriculture, services, and trade sectors, reflecting significant geographic and socio-cultural heterogeneity. Figure 1 presents the study area along with the official GRDP distribution of East Java Province in 2022.

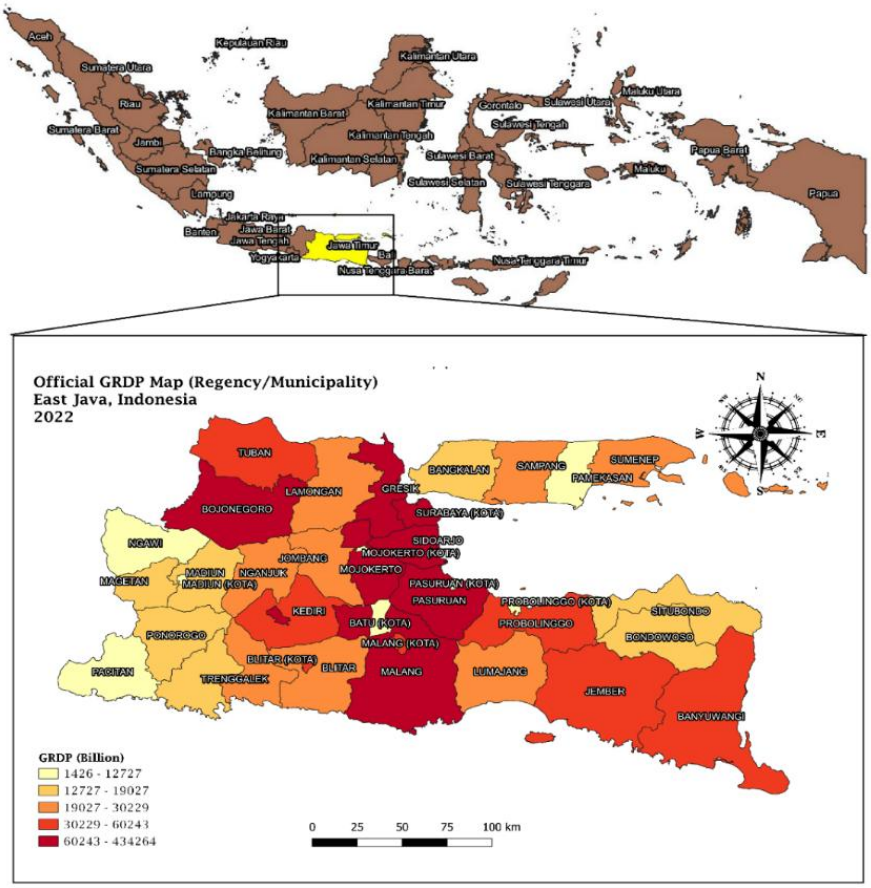

Figure 1. The official GRDP at constant prices mapping in East Java, Indonesia in 2022

2.2. Big Data

Big data has become a crucial concept in both industry and academia, encompassing various meanings. Big data as an information asset characterized by volume, velocity, and variety (the 3Vs), requiring specific technologies and analytical methods to derive value [29]. Its application is expanding rapidly across sectors such as industry, environment, and economic development. A specialized form, geospatial big data, refers to spatially referenced data, including land characteristics like land use [30]. This data generally appears in two formats: raster data, comprising images from sources like drones or satellites (e.g., Google Earth), and vector data, consisting of points, lines, and polygons obtained through platforms such as the OpenStreetMap API [30].

2.3. Satellite Imagery

Satellite imagery refers to images captured by sensors mounted on satellites orbiting more than 400 km above Earth's surface [31]. These images result from remote sensing, a process in which sensors record energy emitted or reflected by objects on Earth [31]. Remote sensing involves collecting data from aircraft or satellites to measure and observe terrestrial features [32] and includes all techniques for capturing and processing electromagnetic recordings of the Earth's surface [33]. Its ability to provide consistent, repetitive observations enables the monitoring of both short- and long-term environmental changes caused by natural or human activities [32].

Remote sensing serves as a data source, a science, and a tool simultaneously. As a data source, it records physical properties of distant objects through emitted or reflected energy. As a science, it applies systematic processes such as measurement, data analysis, and interpretation. As a tool, it supports applications ranging from resource inventory to ecological assessment [34].

Remote sensing data are characterized by four key resolutions:

1. Spatial Resolution, the smallest detectable object in imagery, represented by pixel size [31]. Higher resolution (smaller pixels) yields more detailed information [35]. Examples include very high-resolution imagery (Piades, 0.5 m), high (Iconos, 4 m), medium (Landsat, 15–30 m), and low (MODIS, 250–1000 m) [31].
2. Temporal Resolution, the frequency with which a satellite revisits the same location [31]. A shorter revisit time means higher temporal resolution; this depends on orbit type, sensor characteristics, and image swath width [35], [36].
3. Spectral Resolution, the sensor's ability to distinguish between different wavelengths of the electromagnetic spectrum. Higher spectral resolution means more and narrower spectral bands [31].
4. Radiometric Resolution, the number of bits used to represent energy levels per pixel, determining brightness detail. Greater bit depth (e.g., 8-bit, 10-bit, 16-bit) indicates higher radiometric resolution [31], [37].

Satellite imagery is available from various sources such as VIIRS, Sentinel-2, MODIS, Sentinel-5P, and ASTER. Subsequent analyses often utilize imagery from these satellites to explore the relationship between environmental observations and socio-economic indicators, including GRDP formation.

2.4. Visible Infrared Imaging Radiometer Suite (VIIRS)

The Suomi National Polar Partnership (SNPP) is a satellite operated jointly by NASA and NOAA, equipped with the Visible Infrared Imaging Radiometer Suite (VIIRS) as its primary imaging sensor. VIIRS is distinct among meteorological satellite sensors for its ability to capture short-wave infrared, near-infrared, and nighttime visible light emissions [38]. Research utilizing nighttime light data expanded following the release of annual stable light products from the DMPS-OLS by NOAA's Earth Observation Group (EOG). Since the launch of VIIRS in 2011, EOG has provided improved nighttime light datasets featuring monthly updates, higher spatial resolution, and enhanced data quality [39]. These advancements make VIIRS data highly valuable for studies analyzing social and economic activities, particularly those related to economic growth [39][40].

2.5. Sentinel-2 Multispectral Instrument Level-2A

Sentinel-2 is a European satellite equipped with the Multispectral Instrument (MSI) Level-2A, launched in 2015, designed for high-resolution optical imaging of the Earth's surface. Unlike radar satellites that utilize Synthetic Aperture Radar Ground Range Detected (SAR-GRD) technology to capture images through cloud cover [41], Sentinel-2 focuses on multispectral observation. Compared to other medium-resolution satellites, Sentinel-2 offers superior spatial resolution, reaching 10 meters in the B4 (Red), B3 (Green), B2 (Blue), and B8 (Near-Infrared/NIR) bands. The satellite's temporal resolution, its revisit frequency is 10 days for a single satellite and 5 days when two satellites operate together.

Sentinel-2 is a multispectral imaging satellite mission that supports the Copernicus Land Monitoring program, focusing on observing vegetation, land cover, water bodies, inland waterways, and coastal areas. Its combination of sensing bands enables the development of various composite indices, including the Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Normalized Difference Built-Up Index (NDBI). Sentinel-2 offers high spatial resolution of up to 10 meters in the B4 (Red), B3 (Green), B2 (Blue), and B8 (Near Infrared/NIR) bands, with a temporal resolution of 10 days, or 5 days when two satellites are operational.

The NDVI quantifies vegetation greenness by analyzing reflectance differences between the NIR and Red bands, as formulated:

$$NDVI = \frac{NIR_{band\ 8} - RED_{band\ 4}}{NIR_{band\ 8} + RED_{band\ 4}} \quad (1)$$

NDVI reflects vegetation health and ecological changes [42] and has shown a positive correlation with economic growth, particularly with GDP and population [16].

The NDWI distinguishes between water and land surfaces, as it is sensitive to liquid water content in vegetation [43]. Its values range from below 0 (non-water areas) to above 0 (water areas) [44], with the following formula:

$$NDWI = \frac{Green_{band\ 3} - NIR_{band\ 8}}{Green_{band\ 3} + NIR_{band\ 8}} \quad (2)$$

The NDBI identifies built-up or urban areas using the difference between Short-Wave Infrared (SWIR) and NIR reflectance [45]. According to [46], it is formulated as:

$$NDBI = \frac{SWIR_{band\ 11} - NIR_{band\ 8}}{SWIR_{band\ 11} + NIR_{band\ 8}} \quad (3)$$

Higher NDBI values indicate a greater likelihood of urban development [47].

2.6. MODIS

The Moderate Resolution Imaging Spectroradiometer (MODIS) is a sensor onboard the Terra satellite, operating under NASA's Earth Observing System [48]. MODIS–Terra captures Land Surface Temperature (LST) data both day and night, providing valuable insights into Earth's thermal conditions. The satellite orbits polarly at an altitude of 705 km, crossing the equator around 10:30 a.m. local time [48].

LST data are particularly useful for analyzing the Urban Heat Island (UHI) phenomenon [22]. Increases in population density, industrial activity, and infrastructure development significantly influence LST variations. Consequently, areas exhibiting higher LST values typically correspond to regions with intense socio-economic activities and denser populations [22].

2.7. Sentinel-5P

The Sentinel-5 Precursor (Sentinel-5P) is an atmospheric monitoring satellite launched by the European Space Agency (ESA) on October 13, 2017, designed to observe and measure various air pollutants [49]. Sentinel-5P captures data on gases such as carbon monoxide (CO), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2), which are key indicators of air quality and human environmental impact.

Carbon Monoxide (CO)

CO is a crucial atmospheric gas for assessing tropospheric chemistry and serves as a major pollutant in urban environments (Copernicus Sentinel-5P, processed by ESA, 2021). It is primarily generated from biomass burning, fossil fuel combustion, and atmospheric oxidation of methane and hydrocarbons, with significant sources including motor vehicles, forest fires, power plants, and incinerators [50].

Nitrogen Dioxide (NO_2)

NO_2 occurs in both the troposphere and stratosphere, produced by anthropogenic activities (e.g., fossil fuel and biomass combustion) and natural processes such as soil microbial reactions, lightning, and wildfires (Copernicus Sentinel-5P, processed by ESA, 2021). It is one of the major air pollutants in industrial and urban regions worldwide [51], [52]. Areas with high economic and industrial activity generally exhibit elevated NO_2 concentrations.

Sulfur Dioxide (SO_2)

SO_2 emissions are closely linked to population density and energy consumption [53]. High population areas tend to produce more SO_2 pollutants. Key contributors include electricity generation, manufacturing, and mining activities. As economic activity increases, greater energy demand for both residential and industrial purposes results in higher SO_2 emissions into the atmosphere [54].

2.8. The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) is a satellite jointly operated by the Japanese Ministry of Economy, Trade, and Industry (METI) and the U.S. National Aeronautics and Space Administration (NASA). Launched in December 1999, ASTER carries 14 imaging sensors capable of capturing thermal infrared, shortwave infrared, and visible-near infrared (VNIR) spectral data [55].

ASTER is notable for its ability to generate Digital Elevation Models (DEMs) three-dimensional representations of the Earth's surface that exclude above-ground objects such as vegetation and buildings [56]. DEMs are widely used in geographic information systems (GIS) and serve as the foundation for creating digital maps. Variations in elevation influence economic activity patterns: higher elevations are typically dominated by agricultural activities, while lower-lying areas tend to support more industrial development.

2.9. Point of Interest (POI)

Point of Interest (POI) refers to specific locations that attract frequent human activity, such as restaurants, convenience stores, transportation hubs, parks, cafés, and tourist attractions, places that provide value or utility to visitors [57]. With the widespread use of mobile devices, service providers and technology companies increasingly analyze user mobility data with high geospatial-temporal resolution to identify and map nearby POIs [57]. Common POI data sources include OpenStreetMap (OSM), Google Maps, HERE Maps, and OneMap. In this study, POI data are obtained from OpenStreetMap (OSM).

2.10. Robust Linear Regression

The modeling used is robust regression modeling that can address the problem of heteroskedasty, and the image (oulier) on the data. Robust regression has many kinds of methods in terms of dealing with the problems of various classical regression assumptions such as M-Regression, S-Regresion, and MM- Regression. In this study, the MM-Regression method will be used because this method combines between M- Regression and S-Regresion so that it is resistant and has a higher efficiency [58]. Robust regression will not make the modeling error following the normal distribution because it has a different estimate method based on the weighing of the data so that the outlier will give a small weight to the parameter, but produce a level of accuracy generated better than the Ordinary Least Square method. (OLS) [59]. The MM-Estimation procedure is a procedure by combining S-Estimation and M-Estimation with the initial step of searching for the Estimator S so that a breakdown point is obtained to minimize the residual in the estimator M which is used to set parameters on its regression so that its efficiency is high and usually uses Tukey Bisquare. Based on [60], the equation form of the estimate method Method of Moment (MM) is as follows.

$$\sum_{i=1}^{n} p_i'(u_i)X_{ij} = 0 \tag{4}$$

$$\sum_{i=1}^{n} p_i'\left(\frac{Y_i - \sum_{j=0}^{k} X_{ij}\hat{\beta}_j}{S_{MM}}\right)X_{ij} = 0 \tag{5}$$

where s_{MM} is the standard deviation derived from S-estimation and ρ is Tukey's biweight function formulated as follows:

$$\rho(u_i) = \frac{u_i^2}{2} - \frac{u_i^4}{2c^2} + \frac{u_i^6}{6c^2}, -c \le u_i \le c \text{ dan}$$
$$\rho(u_i) = \frac{c^2}{6}, u_i < -c \text{ atau } u_i > c \tag{6}$$

calculating the weighted value derived from Tukey Bisquare in the following equation.

$$w_i = [1 - (\frac{u_i}{c})^2]^2, u_i \leq c$$
$$w_i = 0,\ u_i > c \tag{7}$$

where $u_i = \frac{e_i}{S_{MM}}$ and $c = 4.865$
The MM estimate method uses WLS (Weighted Least Square) to estimate the regression parameters in the form of the following equation.

$$\hat{\beta}_i^{(n)} = (X^T W^{(n)} X)\ (X^T W^{(n)} Y) \tag{8}$$

where n is the value of the iteration performed and this iterative process is performed so that the weighing value becomes convergent or $\hat{\beta}_i^{(n+1)} - \hat{\beta}_i^{(n)} \approx 0$.

2.11. Data Transformation

The transformation used in this study is the Yeo-Johnson power transformation, which is a family of Box-Cox transformations and can transform positive and negative data. The Yeo-Johnson power transformation is well used in handling variables that do not have the same units in all ranges by making the distribution of these variables like a Gaussian distribution or more normally distributed [61]. Data transformation in this study is defined as follows:

$$y_\lambda(x) \begin{cases} \frac{(1+x)^\lambda - 1}{\lambda}, \lambda \neq 0\ dan\ x \geq 0 \\ log(1+x), \lambda = 0\ dan\ x \geq 0 \\ -\frac{(1-x)^{2-\lambda}}{2-\lambda}, \lambda \neq 2\ dan\ x < 0 \\ -log(1-x), \lambda = 2\ dan\ x < 0 \end{cases} \tag{9}$$

where x is the value of the variable or data input and λ is the parameter value estimated using the Maximum Likelihood technique with the assumption that these variables follow a normal distribution. A linear relationship will be generated by families with parameter $\lambda = 1$, then the transformation will make the right tail of the distribution thicker or denser at 1 while widening the left tail of the distribution to make the distribution right-skewed (skewed) towards a symmetrical distribution. In λ > 1, the distribution that is left skewed becomes a more symmetrical distribution as well.

2.12. Corelation Analysis and Selection of Important Variabels

The weighted sum model with the linear model is a weighting calculation to build the RSGI index used in this study. Correlation analysis and selection of important variables need to be carried out to ensure that the variables used can linearly represent the GRDP mapping in East Java Province. The numerical relationship of integrated multi-source satellite imagery and POI data with regional GRDP data at the sub-district level is carried out using correlation analysis. Correlation analysis was carried out by using the correlation coefficient of Pearson and Spearman Rank with a p-value that showed significant results at a significance level of 5% ($\alpha = 0.05$). The following equation is a formula for obtaining the value of the Pearson correlation coefficient.

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{\sqrt{(n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}\sqrt{(n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}} \tag{10}$$

The value of r_{xy} is the correlation value between x_i and y_i, and n is the number of observations. Meanwhile, the Spearman correlation value is calculated by changing the observed value to the ranking value as in the following equation.

$$r_s = 1 - \frac{6\ \sum_{i=1}^{n} d_i^2}{n^3 - n} \tag{11}$$

The value of the correlation coefficient is between 0 and 1, with the direction of the relationship indicated by a positive or negative sign. Table 1 below shows the guidelines for interpreting the results of the correlation coefficient used in this study [62].

Table 1. Correlation coffecient interpretation

Correlation Coefficient	Interpretation
$0.00 \leq \lvert r \rvert \leq 0.199$	Very Weak
$0.20 \leq \lvert r \rvert \leq 0.399$	Weak
$0.40 \leq \lvert r \rvert \leq 0.599$	Moderate
$0.60 \leq \lvert r \rvert \leq 0.799$	Strong
$0.80 \leq \lvert r \rvert \leq 1.000$	Very Strong

Furthermore, a correlation significance test was performed to determine whether the correlation coefficient obtained was statistically significant with a significance level of 5% ($\alpha = 0.05$). The null hypothesis used in this study is that there is no correlation or relationship between the two variables x_i and y_i, compared with the alternative hypothesis, which assumes that there is a relationship between the variables x_i and y_i. Based on the results of the correlation analysis, a statistically significant correlation with the GRDP variable in East Java will be selected at the sub-district level and will then be used for the construction of the RSGI. Thus, RSGI is made based on variables originating from satellite imagery big data and other geospatial data, namely POI, which linearly has a significant relationship or correlation to estimate GRDP in East Java Province down to the sub-district granular level.

2.13. Weighted Sum Model

The relative spatial GDP index (RSGI) is calculated by integrating satellite imagery and geospatial variables that represent the economic activity in East Java Province by calculating the weighted sum model. Several previous studies have used the index building method in their research, and some of them have used geospatial data [40], [63]–[66]. The variables used in the index construction include NTL Total, POI density, POI distance, NDVI, NDBI, NDWI, LST Daytime, LST Nighttime DEM, and atmospheric pollutants (NO_2, SO_2, CO). The following formula is used to build RSGI with a weighted sum model:

$$\begin{aligned} RSGI\ =\ & w_1.NTL + w_2.POI_{density} + w_3.POI_{distance} + w_4.NDVI + w_5.NDWI \\ & + w_6.NDBI + w_7.LST_{daytime} + w_8.LST_{nighttime} + w_9\ .DEM + w_{10}.CO \\ & + w_{11}.NO_2 + w_{12}.SO_2 \end{aligned} \tag{12}$$

where p is the number of variables used in building the index, w is the weight used, and x_i is the value of each observation.

In this study, the RSGI is not a single model instead, it is constructed in four distinct variations based on different weighting approach. They are equal weight sum (RSGI EWS), pearson correlation (RSGI Pearson Weight), spearman correlation (RSGI Spearman Weight), and principal component analysis loadings (RSGI PCA). First, an equal-weighted correlation method was applied, where all variables were assigned identical weights, influenced only by the direction of their correlation with GRDP. Second, weights were determined based on the Pearson correlation coefficient, using only variables that showed significant correlations with GRDP in East Java. Third, the Spearman Rank correlation coefficient was used, assigning higher weights to variables with stronger correlations, representing GRDP more accurately. Fourth, weights were derived from Principal Component Analysis (PCA), utilizing components that contributed cumulatively significant variance above 70%. Several previous studies have also employed PCA for index construction in socio-economic research, who developed the Regional Digital Development Index (RDDI) for Indonesia [67]; who created a Spatial

Deprivation Index [68]; and who mapped the Relative Spatial Poverty Index (RSPI) in East Java, showing a strong correlation with poverty levels [65].

2.14. Evalution

To find out how far the RSGI has come in estimating the GRDP of an area in East Java, an evaluation measure is used. In this study, the evaluation used is a numerical and descriptive evaluation. A descriptive evaluation measure is carried out by visually identifying the RSGI results obtained. Then, numerical evaluation is used to numerically measure the accuracy of the results between the results obtained and the ground truth results, where in this case the Pearson, Spearman Rank, RMSE, MAE, MAPE, and R^2correlation measures are used in the robust linear regression model at the sub-district level. RMSE, MAE, MAPE, and R^2 values are calculated using the following formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{13}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{14}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| x\ 100\% \tag{15}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \underline{y}_i)^2} \tag{16}$$

2.15. Analysis Method

This study utilizes multi-source satellite imagery and Point of Interest (POI) data from OpenStreetMap (OSM), collected between January 1 and December 31, 2022. The satellite data were collected and processed using the Google Earth Engine (GEE) platform. The variables used are summarized in Table 1. The target variable, Gross Regional Domestic Product (GRDP) at constant prices (ADHK), was chosen because it reflects the real economic growth rate. Since official GRDP data are only available at the regency or city level, GRDP for the 666 sub-districts was estimated using the Broad Area Ratio Estimation (BARE) method, a Small Area Estimation (SAE) technique that employs population data as a proportional auxiliary variable.

The remote sensing data used in this study consist of multi-source satellite imagery and geospatial big data. The satellite imagery includes Nighttime Light (NTL) intensity from NOAA-VIIRS; Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Normalized Difference Built-Up Index (NDBI) from Sentinel Multispectral Level 2A; Daytime and Nighttime Land Surface Temperature (LST) from MODIS; Nitrogen Dioxide (NO_2), Carbon Monoxide (CO), and Sulfure Dioxide (SO_2) from Sentinel-5P; as well as Digital Elevation Model (DEM) data from ASTER. These variables serve as GRDP estimation indicators based on relevant literature, previous studies, and image characteristics identified for East Java Province. The data were collected and processed through Google Earth Engine (GEE), a cloud-based platform for managing and analyzing Earth observation data, offering high-resolution, freely accessible imagery suitable for advanced processing such as classification [69]. Meanwhile, POI data represent another form of geospatial big data collected from OSM to complement the satellite-based analysis.

POI data describe the accessibility and activity level of an area. In this study, more than 17,000 POI points were identified across East Java Province, representing public facilities categorized into sectors such as education, health, economy, and tourism. The highest satellite spatial resolution is provided by Sentinel Multispectral Level 2A (10 m), while the lowest is Sentinel-5P (1113.2 m). To address varying resolutions, GRDP mapping was constructed through the Relative Spatial GDP Index (RSGI) with a resolution of 1 km × 1 km, in which each grid cell aggregates extracted features from all variables at the sub-district level. All geospatial data sources used in this study, including POI data from OSM and multisource remote sensing data underwent a validation and quality-checking procedure prior to modeling. The POI dataset was refined through classification cleaning, removal of duplicated or irrelevant entries, and spatial consistency checks to ensure its representativeness for economic activity. In addition, the completeness and distribution of POI categories were examined to confirm that

economically relevant facilities such as commercial, industrial, and service-related points were adequately captured within each sub-district. Then, the satellite-derived variables were obtained from products that include sensor calibration and atmospheric correction. Further validation was implicitly ensured through the standardized pre-processing workflow applied to each satellite dataset, which included cloud masking, median compositing, band selection, and pollution-product selection as appropriate for each sensor. These procedures reduce atmospheric noise, remove anomalous observations, and enhance the reliability of the extracted indicators for representing underlying economic conditions. These steps ensured that each data source met the standards of methodological appropriateness and was fit for use in constructing the RSGI. Additionally, official population data for 2022 from Statistics Indonesia (BPS) at the regency, municipality, and sub-district levels were used to develop a GRDP estimation model using the SAE approach with the BARE method, an approach widely applied in previous studies [70]. The GRDP at constant prices (ADHK) was employed as it reflects the real rate of economic growth across sectors and over time [28].

This research proposes a method to overcome the limitations of conventional surveys for GRDP calculation, which are often costly, slow, and lack granularity. The proposed solution utilizes multi-source satellite imagery and OpenStreetMap (OSM) Point of Interest (POI) data to estimate GRDP at the sub-district and district/city levels using a robust linear regression model. The overall research framework is depicted in Figure 2.

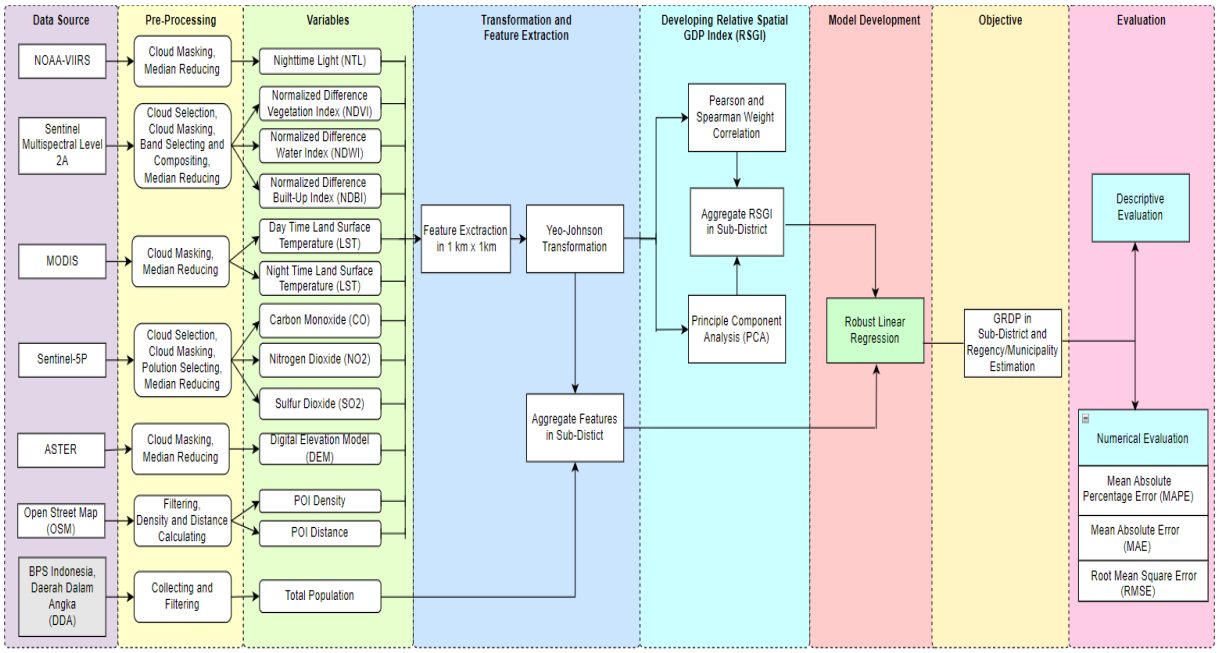

Figure 2. The research framework

The methodology is centered on the creation of a Relative Spatial GDP Index (RSGI). This process begins with preprocessing and feature extraction from a diverse set of variables: NTL, NDVI, NDWI, NDBI, NO_2, CO, SO_2, daytime and nighttime LST, POI Density, POI Distance, and DEM. These features are aggregated into a 1 km x 1 km grid, transformed using the Yeo-Johnson method, and then combined to construct the RSGI.

This index serves as the primary predictor in a robust linear regression model to estimate GRDP. The performance and accuracy of the model are evaluated using several metrics: the correlation coefficient, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (). Finally, the GRDP estimation results are presented as visualized maps to facilitate interpretation and support stakeholder decision-making.

3. Results and Discussion

3.1. Correlation Analysis and Principal Component Analysis (PCA)

In this study, correlation analysis was applied to determine the direction and strength of the relationship between geospatial variables and GRDP in East Java. Due to the limited availability of official GRDP data at the 1 km × 1 km grid level, the analysis was conducted at the district or city level (38 observations) and at the sub-district level (666 observations) using GRDP estimates derived from the BARE method. To obtain more accurate insights, both Pearson and Spearman Rank correlation tests were employed, with p-values used to assess statistical significance at $\alpha = 5\%$ and correlation coefficients indicating the direction and closeness of relationships. Following the interpretation approach of [62], the Pearson test results revealed that all variables, except Nighttime LST, were significantly correlated (p-value < 0.05) with GRDP at the sub-district level, while the Spearman Rank test showed significant correlations (p-value < 0.05) for all variables. The results of the correlation analysis are explained in Table 2.

Table 2. Results of correlation analysis of geospatial data variabels on GRDP in East Java

Variable	Pearson Correlation Test		Spearman Rank Test		Direction	Closeness	Statistically Significant
	Correlation Coefficient	p-value	Correlation Coefficient	p-value			
NTL	0.6031	2.2×10^{-16}	0.5677	2.2×10^{-16}	Positive	Strong & Moderate	Significant
NDVI	-0.3645	2.2×10^{-16}	-0.3019	1.8×10^{-15}	Negative	Weak	Significant
NDWI	0.3506	2.2×10^{-16}	0.3014	2.1×10^{-15}	Positive	Weak	Significant
NDBI	0.3840	2.2×10^{-16}	0.326	2.2×10^{-16}	Positive	Weak	Significant
Day LST	0.2395	4.1×10^{-10}	0.2699	1.5×10^{-12}	Positive	Weak	Significant
Night LST	0.0627	1.1×10^{-1}	0.1116	4.0×10^{-3}	Positive	Very Weak	Not Significant
CO	0.2099	4.8×10^{-8}	0.2553	2.5×10^{-11}	Positive	Weak	Significant
NO_2	0.3753	2.2×10^{-16}	0.3389	2.2×10^{-16}	Positive	Weak	Significant
SO_2	0.2845	7.8×10^{-14}	0.2947	9.04×10^{-15}	Positive	Weak	Significant
POI *Density*	0.4974	2.2×10^{-16}	0.4288	2.2×10^{-16}	Positive	Moderate	Significant
POI *Distance*	-0.4017	2.2×10^{-16}	-0.3853	2.2×10^{-16}	Negative	Moderate	Significant
DEM	0.2845	7.8×10^{-14}	0.2947	9.04×10^{-15}	Positive	Weak	Significant

The PCA method was applied to reduce data dimensions and extract principal components that capture the highest variance proportions from the original data. Based on [71], components contributing a cumulative variance above 70% were selected. As shown in Table 3 and Figure 3, three components were chosen, accounting for 78.003% of the total variance. Component 1 explains 38.821% of the variance, dominated by high loadings from NTL, NDBI, NO_2, and CO variables. Component 2 contributes 22.6% (cumulative 61.4252%), with SO_2 and DEM showing the highest loadings. Component 3 adds 16.58%, bringing the cumulative variance to 78.003%, with dominant loadings from NTL, NDVI, and NDWI. These selected components serve as weighted coefficients in constructing the RSGI using the weighted sum method based on the obtained loading values.

Table 3. Loadings value (eigen vector) of each main component of PCA

Variable	Component											
	1	2	3	4	5	6	7	8	9	10	11	12
NTL	0.308	0.100	0.708	-0.164	0.023	0.026	-0.114	0.800	0.454	-0.016	-0.027	0.0000
NDVI	-0.171	-0.117	-0.555	-0.142	-0.222	0.065	0.143	0.155	-0.009	-0.015	0.727	1.8×10^{-15}
NDWI	0.137	0.130	0.580	0.172	0.257	0.039	-0.238	-0.094	-0.014	-0.050	0.678	1.6×10^{-15}
NDBI	0.434	-0.049	0.084	-0.047	-0.045	0.502	0.677	-0.146	0.157	0.178	0.084	-3.4×10^{-16}
NO2	0.486	0.004	-0.136	0.553	-0.103	0.624	0.184	-0.062	0.016	-0.006	0.008	-1.2×10^{-16}
CO	0.395	-0.244	-0.286	0.308	-0.150	0.506	-0.508	0.053	-0.245	-0.073	-0.020	4.3×10^{-16}
SO2	-0.017	0.621	-0.242	0.128	0.133	0.134	-0.009	-0.022	-0.002	0.041	-0.002	-7.1×10^{-16}
Day LST	0.291	-0.061	-0.229	-0.339	0.481	0.069	0.021	-0.209	0.050	-0.677	-0.015	3.1×10^{-15}
Night LST	0.197	-0.163	-0.225	-0.237	0.503	0.146	-0.205	-0.128	0.020	0.701	0.025	-3.6×10^{-15}
POI Density	0.288	0.208	0.138	-0.393	-0.143	0.142	0.097	0.222	-0.772	0.044	-0.002	2.3×10^{-16}
POI Distance	-0.260	-0.224	-0.035	0.400	0.553	0.109	0.324	0.431	-0.327	-0.045	-0.029	5.4×10^{-16}
DEM	-0.017	0.621	-0.232	0.128	0.133	0.134	-0.009	-0.022	-0.002	0.041	-0.002	-1.2×10^{-16}

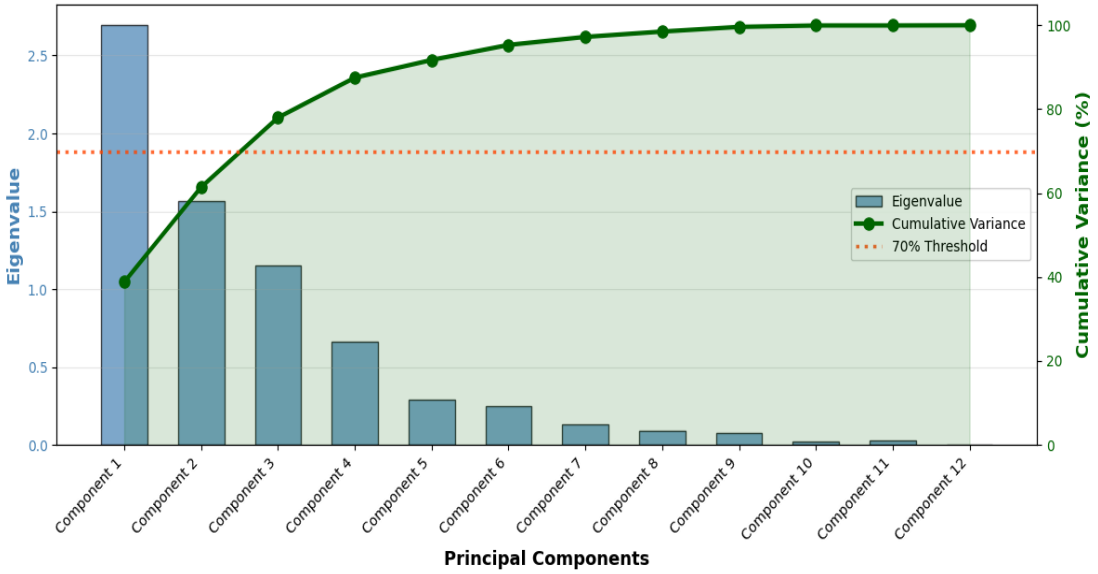

Figure 3. Principal component selection and cumulative variance contribution

3.2. Calculation of the RSGI (Relative Spatial GDP Index)

To provide an initial understanding of the linear relationships among the key geospatial indicators and the dependent variable, bivariate analyses were conducted, focusing on the three parameters exhibiting the strongest association with the logarithm of GRDP (Ln GRPD), they are NTL, POI Density, and POI Distance. The visual representation of these relationships is presented through scatter plots in Figure 4. The first plot illustrates the moderately strong positive relationship between NTL Total and Ln GRDP, indicate that areas with higher aggregate night-time light intensity consistently exhibit higher economic output, affirming the known strong link between energy consumption, urbanization, and economic activity.The second plot examines the association between POI Density and Ln GRDP, also demonstrates a positive relationship and this finding reflects the principle that regions with a denser concentration of economic activities, services, and infrastructure (as captured by the POI count per unit area) tend to be associated with significantly higher GRDP values. Conversely, the third plot focuses on POI Distance, which shows a clear negative association with Ln GRDP, and this pattern confirms the expectation that greater average distance from economic centers is typically linked to a decrease in overall economic activity. Overall, these initial visual patterns and correlation values, strongly reinforce the relevance and potential of these geospatial indicators in effectively capturing and explaining the observed variations in regional economic intensity across the study area.

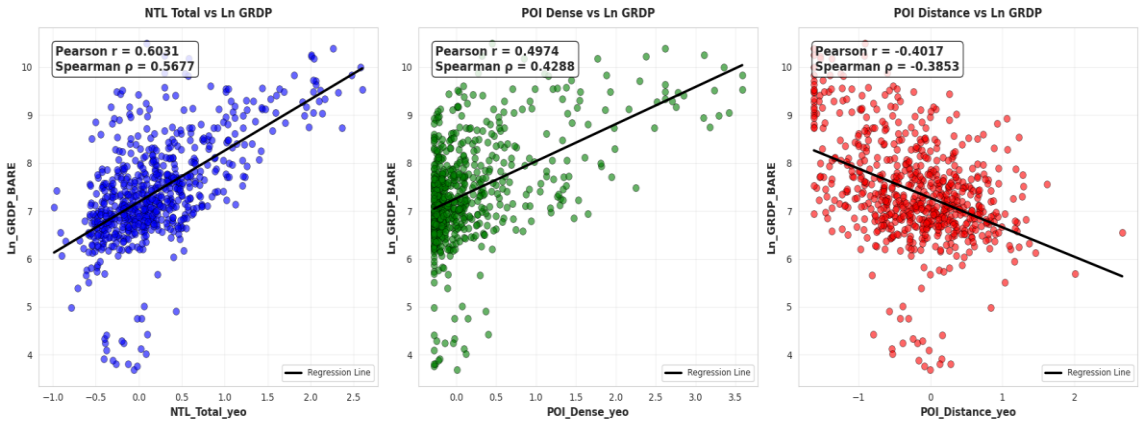

Figure 4. Correlation analysis between socioeconomic indicators (NTL, POI density, and POI distance) and regional economic output in East Java districts

In building a linear regression model, selection is carried out using the Pearson correlation test with a moderate to strong relationship. The results of the calculations are shown in Table 4, sorted from the highest value. The Relative Spatial GDP Index (RSGI) was calculated using the weighted sum model method, incorporating geospatial variables that showed significant correlations with GRDP in East Java Province, namely NTL, NDVI, NDWI, NDBI, NO_2, CO, SO_2, Daytime LST, Nighttime LST, POI Density, POI Distance, and DEM. Two weighting strategies were applied: an equal weighted sum, which assigns identical weights to all variables based solely on the direction of their significant correlations, and four specific weighting approaches, equal weights based on correlation direction (W_1), Pearson correlation coefficients (W_2), Spearman Rank correlation coefficients (W_3), and PCA-based weights derived from the first three components (W_4), which together account for 78% of cumulative variance. The results of these weight calculations are presented in the following table.

Table 4. RSGI weight calculation results

Variable	Equal Weight	Pearson Weight	Spearman Weight	PCA Weight
NTL	+1	0.6031	0.5677	1.116
NDVI	-1	-0.3645	-0.3019	-0.843
NDWI	+1	0.3506	0.3014	0.847
NDBI	+1	0.3840	0.3260	0.469
NO2	+1	0.3753	0.3389	0.354
CO	+1	0.2099	0.2553	-0.135
SO2	+1	0.2845	0.2947	0.309
Daytime LST	+1	0.2395	0.2699	0.001
Nighttime LST	+1	Not Significant	0.1116	-0.191
POI Density	+1	0.4974	0.4288	0.634
POI Distance	-1	-0.4017	-0.3853	-0.519
DEM	+1	0.2845	0.2947	0.372

Using four different weighting approaches, the weighting was applied to each 1 km × 1 km grid cell to generate a relative spatial GDP (RSGI) map, as shown in Figure 5. The mapping process employed min–max scaling to normalize RSGI values within a 0–1 range for easier interpretation. The RSGI values were classified into six categories using the equal interval method, ensuring each class spans an identical range, and the color gradient signifies an increasing index value. Very Low RSGI areas, ranging from 0 to 0.17, are depicted in black, representing the lowest combined socio-economic, demographic, and physical-geographic activity. This is followed by Low RSGI (0.17–0.33) in dark purple, Moderate Low RSGI (0.33–0.5) in purple, and Moderate High RSGI (0.5–0.67) in dark orange. The higher end of the spectrum includes High RSGI (0.67–0.83), shown in orange, and finally, Very High RSGI (0.83–1), depicted in bright cream, which indicates the highest concentration of the combined index variables.

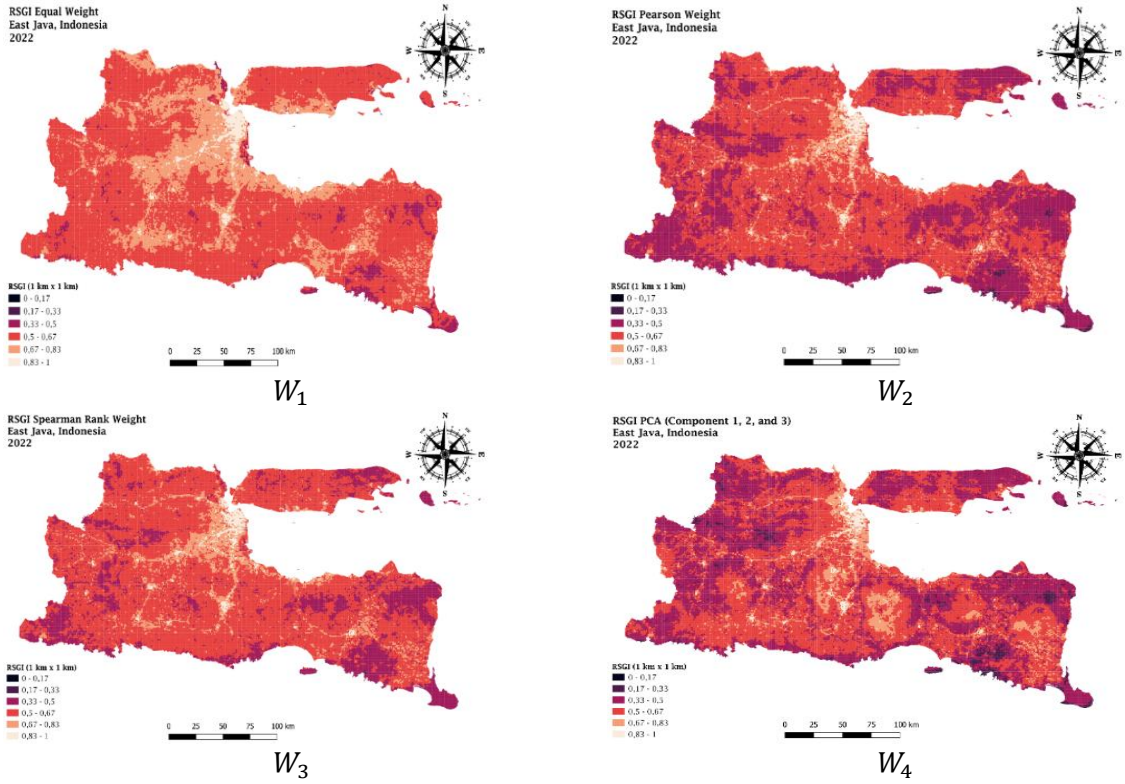

Figure 5. RSGI with equal weight sum method visualization

The resulting of four different weights for RSGI maps consistently indicate a stark spatial dichotomy. Higher index values, predominantly characterized by orange and bright cream colors are primarily concentrated in major urban centers and their expanding peri-urban hinterlands. The most significant concentration of Very High RSGI forms a dense, continuous cluster centered around the Gerbangkertosusila metropolitan region, encompassing the core area of Kota Surabaya and its neighboring highly industrialized municipalities and Kota Malang. This expansive, high-value stretch directly reflects the region's status as the uncontested economic and administrative engine of East Java, fueled by high population density, intensive industrial activities, and commercial infrastructure development. Furthermore, other key regional urban clusters, including Jember, Blitar, and Kota Madiun, also exhibit prominent, localized pockets of high RSGI, radiating outward from their municipal boundaries and serving as vital secondary nodes for regional socio-economic growth.

In sharp contrast, lower RSGI values, represented by dark purple and black colors, dominate the less urbanized and peripheral areas, indicating significant disparities. The lowest RSGI values are widespread across the central and eastern parts of Madura Island and are strongly correlated with structural factors such as lower levels of industrial investment, limited access to advanced education and healthcare, and a predominant reliance on traditional economic sectors like agriculture and fisheries. The low values in the south are primarily linked to the difficult topography of the Southern Mountains, which limits agricultural productivity, hinders transportation network development, and restricts high-density economic and social infrastructure. Similarly, low RSGI values in several eastern mountainous areas are associated with forest dominance, conservation land use, and lower population carrying capacity. This spatial dichotomy underscores the geographically driven relationship between urbanization, infrastructure development, and socio-economic conditions in East Java Province.

3.3. Robust Linear Regression Modelling

The study employed robust regression modeling to address issues of heteroscedasticity and outliers. Two models were developed using GRDP (natural logarithm) from the SAE BARE method as the dependent variable (Table 5). Model 1 included only the RSGI as the independent variable, while Model 2 added the population variable (natural logarithm), denoted as *P*.

In Model 1, four robust regression equations were estimated, each yielding significant results for both the intercept and RSGI (p-value < 0.05). The model using RSGI with Pearson weighting achieved an $R^2$of 0.4755, indicating that RSGI explains approximately 47.55% of the GRDP variance. The Breusch–Pagan test confirmed no heteroscedasticity (p-value > 0.05).

In Model 2, both RSGI and population variables were significant (p-value < 0.05). The highest Adjusted R^2 value of 0.7658 was obtained using the RSGI Pearson Weight combined with population, meaning 76.58% of the GRDP variance was explained, while 23.42% remained unexplained. The Breusch–Pagan test again indicated no heteroscedasticity (p-value > 0.05), and the Variance Inflation Factor (VIF) values were all below 10, confirming the absence of multicollinearity [72].

Table 5. Robust Regression modelling results

Information	RSGI Equal Weighted	RSGI Pearson Weighted	RSGI Spearman Weighted	RSGI PCA Weighted
Model 1	$7.212 + 0.116\ RSGI$	$7.209 + 0.334\ RSGI$	$7.207 + 0.349\ RSGI$	$7.282 + 0.263\ RSGI$
Student-t Test (p-value)				
Intercept	2×10^{-16}	2×10^{-16}	2×10^{-16}	2×10^{-16}
RSGI	2×10^{-16}	2×10^{-16}	2×10^{-16}	2×10^{-16}
R^2	0.4262	0.4755	0.4611	0.4392
Breusch Pagan Test (p-value)	0.2904	0.1587	0.1885	0.1819
Model 2	$-4.071 + 0.074\ RSGI$ $1.035\ Ln\ (P)$	$-3.589 + 0.221\ RSGI$ $0.991\ Ln\ (P)$	$-3.071 + 0.229\ RSGI)$ $0.999\ Ln\ (P)$	$-4.271 + 0.1719\ RSGI$ $1.056\ Ln\ (P)$
Student-t Test (p-value)				
Intercept	2×10^{-16}	2.44×10^{-15}	2.44×10^{-15}	2×10^{-16}
RSGI	2×10^{-16}	2×10^{-16}	2×10^{-16}	2×10^{-16}
Ln (Population)	2×10^{-16}	2×10^{-16}	2×10^{-16}	2×10^{-16}
Adjusted R^2	0.7445	0.7658	0.7565	0.7618
Breusch Pagan Test (p-value)	0.2237	0.1726	0.193	0.1032
VIF				
RSGI	1.3236	1.3706	1.3599	1.3362
Ln (Population)	1.3236	1.3706	1.3599	1.3362

3.4. RSGI Numerical and Descriptive Evaluation Results

The first numerical evaluation aimed to assess the closeness between RSGI values and GRDP. Since official GRDP data are only available at the district or city level, numerical evaluation at the 1 km × 1 km grid level was not feasible. To achieve a more accurate assessment at a finer scale, RSGI values were aggregated by sub-district, resulting in 666 observations. The relationship and degree of association between RSGI and GRDP were then evaluated using Pearson and Spearman Rank correlation analyses. The results of these analyses are presented in Table 6.

Table 6. RSGI correlation analysis with GRDP (BARE) in sub-district

RSGI	Pearson Correlation Test		Spearman Rank Test		Direction	Closeness	Statistically Significant
	Correlation Coefficient	p-value	Correlation Coefficient	p-value			
RSGI *Equal Weighted*	0.58429	2.2×10^{-16}	0.54944	2.2×10^{-16}	Positive	Moderate	Significant
RSGI *Pearson Weighted*	0.61250	2.2×10^{-16}	0.58458	2.2×10^{-16}	Positive	Strong, Moderate	Significant
RSGI *Spearman Weighted*	0.60601	2.2×10^{-16}	0.57782	2.2×10^{-16}	Positive	Strong, Moderate	Significant
RSGI PCA *Weighted*	0.59622	2.2×10^{-16}	0.54167	2.2×10^{-16}	Positive	Moderate	Significant

Based on Table 6, all RSGI values show significant correlations in both Pearson and Spearman Rank tests with sub-district level GRDP data. All correlations are positive, indicating that higher RSGI values tend to align with higher GRDP values. According to the interpretation by [62], the Pearson correlation results reveal that RSGI weighted by correlation coefficients exhibit strong relationships with GRDP, while other RSGI variants show moderate correlations. Meanwhile, the Spearman Rank test indicates moderate correlations across all RSGI values. The second numerical evaluation assessed

the predictive accuracy of Model 1 and Model 2 using robust regression (Table 7). Model 1 achieved the lowest RMSE, MAE, and MAPE values when using the RSGI Pearson Weighted variable as the independent variable. In contrast, Model 2, which included population and RSGI PCA Weighted as independent variables, yielded the smallest RMSE, MAE, and MAPE values 0.730, 0.482, and 7.000%, respectively. Overall, Model 2 demonstrated better predictive performance than Model 1, making it the most suitable model for estimating and predicting sub-district level GRDP in East Java Province.

Table 7. Robust Regression evaluation

Information	RMSE	MAE	MAPE
Model 1			
RSGI Equal Weighted	0.87611	0.63364	9.30421%
RSGI Pearson Weighted	**0.85345**	**0.61129**	**9.00263%**
RSGI Spearman Weighted	0.85883	0.61742	9.08007%
RSGI PCA Weighted	0.86731	0.62695	9.24068%
Model 2			
RSGI Equal Weighted	0.74442	0.49498	7.19380%
RSGI Pearson Weighted	0.73358	0.47902	6.98911%
RSGI Spearman Weighted	0.73831	0.48622	7.07949%
RSGI PCA Weighted	**0.73047**	**0.48185**	**7.00055%**

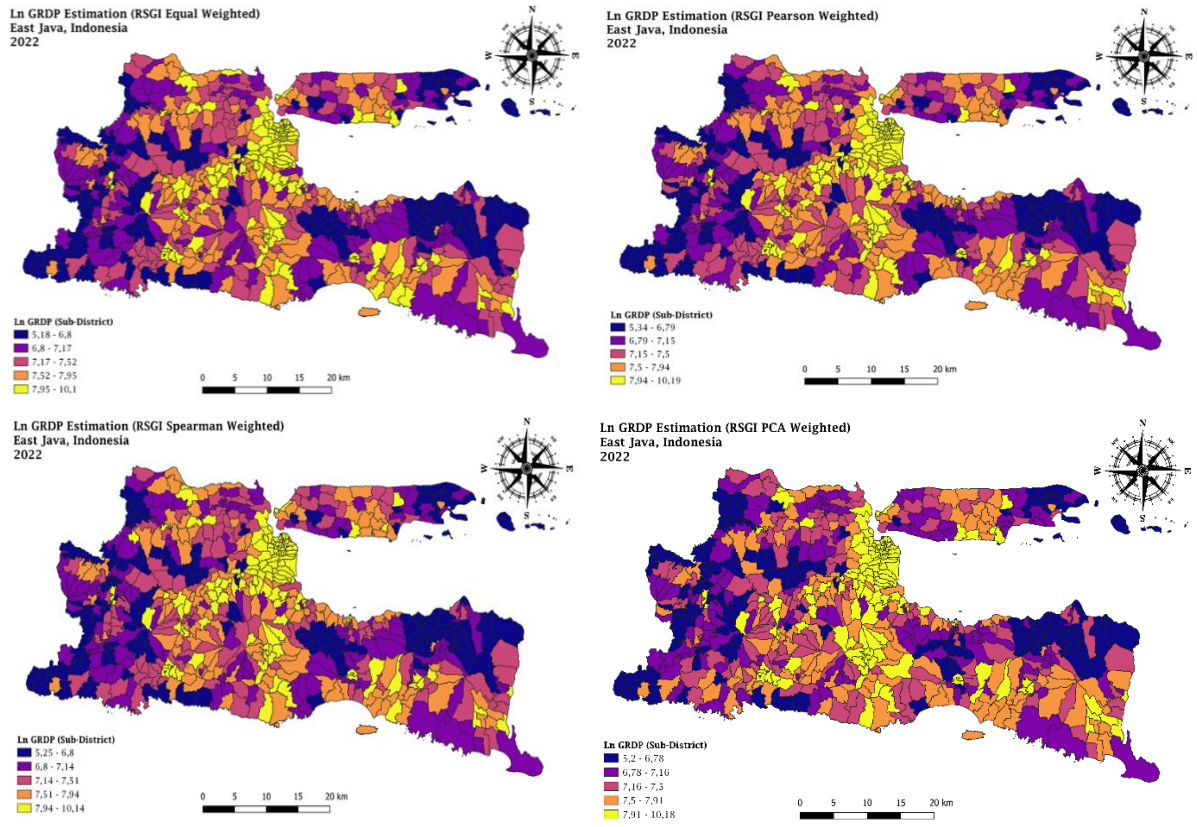

Figure 6. Map of GRDP estimated results in sub-district, East Java 2022 with Robust Regression.

The GRDP estimation results for each model using the robust regression method in Model 2 were visualized to facilitate interpretation, as shown in Figure 6. The sub-district-level GRDP estimates were mapped based on the best-performing model from the numerical evaluations. As illustrated in Figure 7, the model using RSGI PCA and population as independent variables produced the most accurate results, with RMSE, MAE, and MAPE values of 0.73047, 0.48185, and 7.00055%, respectively. Mapping was performed by classifying the estimated GRDP values into five equal intervals at both sub-district and

district/city levels. The sub-district-level results indicate that high GRDP estimates are concentrated in urban centers such as Surabaya City, Malang City, Madiun City, and Kediri City, as well as in several sub-districts of Malang and Banyuwangi Regencies. When aggregated to the district or city level, the highest estimated GRDP values were found in Surabaya City, Sidoarjo Regency, Batu City, Malang City. Overall, the results reveal that high GRDP estimates are concentrated in the northern and southern regions of East Java Province, while lower estimates are predominantly found in the western areas.

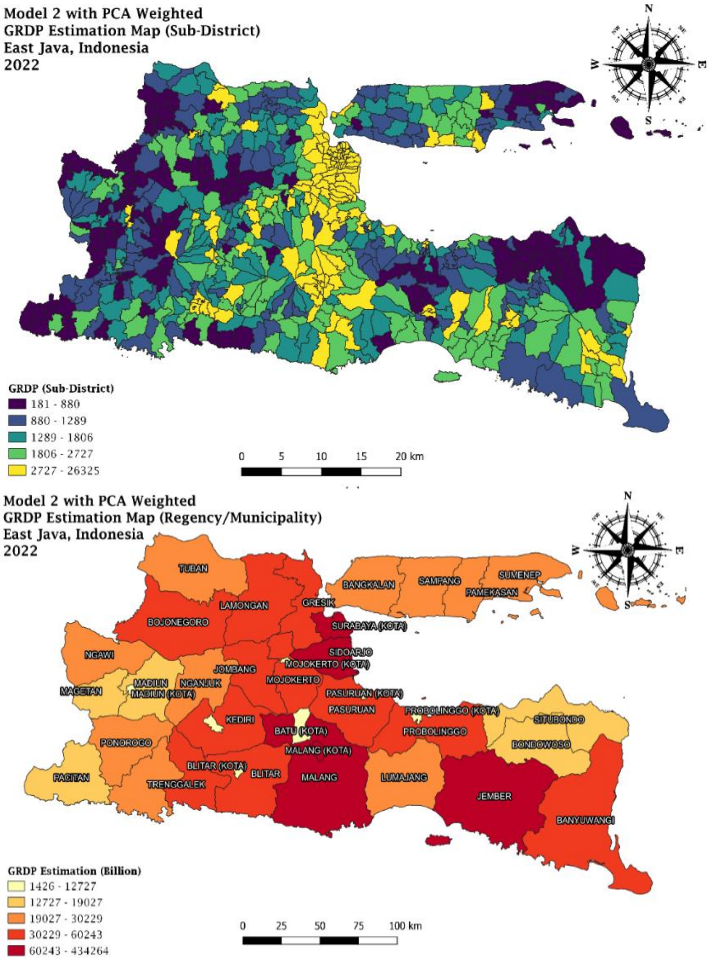

Figure 7. The map of GRDP estimation in regency/municipality level

4. Conclusion

This study introduces a new approach to GRDP mapping in East Java with higher spatial granularity, extending coverage to the sub-district level through the construction of a Relative Spatial GDP Index (RSGI) at a 1 km × 1 km resolution. The method offers a cost-effective and timely update mechanism to support SDG 8, promoting sustained economic growth and monitoring regional development. The research integrates multi-source satellite imagery and Points of Interest (POI) data to build and model the RSGI, estimate GRDP using robust regression, and visualize the resulting spatial distribution.

The developed RSGI demonstrates a strong and significant correlation with sub-district-level GRDP in East Java. Visual analysis indicates that areas with high RSGI values correspond to urban regions with high accessibility, while low RSGI values are concentrated in spatially deprived or less accessible areas. This finding confirms that higher RSGI values are associated with higher GRDP levels.

Two robust regression models were tested, Model 1 using only RSGI and Model 2 incorporating both RSGI and population (log-transformed). Model 2 outperformed Model 1, yielding the best accuracy metrics with RMSE = 0.73047, MAE = 0.48185, and MAPE = 7.00055%, and explaining 76.18% of GRDP variance using the RSGI PCA Weighted variable and population as predictors. The results further reveal that areas with high GRDP estimates are densely populated and well-connected, whereas low GRDP estimates are observed in sparsely populated or undeveloped regions.

Ethics approval

Not required.

Acknowledgments

Not required.

Competing interests

All the authors declare that there are no conflicts of interest.

Funding

This study received no external funding.

Underlying data

This research uses secondary data obtained from Badan Pusat Statistik, Satellite Imagery, and OpenStreetMap.

Credit Authorship

Rifqi Ramadhan: Writing – review & editing, Resources, Methodology, Data Collection, Investigation, Formal analysis, Conceptualization. **I Made Satria Ambara Putra:** Writing – original draft, Investigation, Formal analysis. **Taufiq Agung Kurniawan:** Writing – original draft, Investigation, Formal analysis. **Fitri Kartiasih:** Validation, Supervision, Writing-Reviewing and Editing. **Raden Muaz Munim**: Writing-Reviewing, Validation. **Somethea Buoy**: Writing-Reviewing, Final Draft.

References

[1] Bappenas, "Rancangan teknokratik rencana pembangunan jangka menengah nasional (RPJMN) 2020-2024," *Kementrian Perenc. Pembang. Nas.*, pp. 2015–2019, 2019.

[2] Badan Pusat Statistik, "Pertumbuhan ekonomi indonesia triwulan IV-2022," *Badan Pusat Statistik*, 2023. .

[3] BPS Jawa Timur, "Produk domestik regional bruto provinsi Jawa Timur menurut lapangan usaha 2018-2022," *Badan Pusat Statistik Provinsi Jawa Timur*, 2023. .

[4] Badan Pusat Statistik, "Indikator perumahan dan kesehatan lingkungan," *Badan Pusat Statistik*, 2022. .

[5] Q. Chen *et al.*, "Mapping China's regional economic activity by integrating points-of-interest and remote sensing data with random forest," *Environ. Plan. B Urban Anal. City Sci.*, vol. 48, no. 7, pp. 1876–1894, 2021, doi: 10.1177/2399808320951580.

[6] Y. Zhou, T. Ma, C. Zhou, and T. Xu, "Nighttime light derived assessment of regional Inequality of socioeconomic development in China," pp. 1242–1262, 2015, doi: 10.3390/rs70201242.

[7] Y. Lu and N. C. Coops, "Bright lights, big city: Causal effects of population and GDP on urban brightness," *PLoS One*, vol. 13, no. 7, pp. 1–15, 2018, doi: 10.1371/journal.pone.0199545.

[8] J. A. Pagaduan, "Do higher-quality nighttime lights and net primary productivity predict subnational GDP in developing countries? Evidence from the Philippines," *Asian Econ. J.*, vol. 36, no. 3, pp. 288–317, 2022, doi: 10.1111/asej.12278.

[9] G. C. McCord and M. Rodriguez-Heredia, "Nightlights and subnational economic activity:

estimating departmental GDP in Paraguay," *Remote Sens.*, vol. 14, no. 5, pp. 1–16, 2022, doi: 10.3390/rs14051150.

[10] Y. Gu, Z. Shao, X. Huang, and B. Cai, "GDP forecasting model for China's provinces using nighttime light remote sensing data," *Remote Sens.*, vol. 14, no. 15, 2022, doi: 10.3390/rs14153671.

[11] X. M. Jin, L. Wan, Y. K. Zhang, and M. Schaepman, "Impact of economic growth on vegetation health in China based on GIMMS NDVI," *Int. J. Remote Sens.*, vol. 29, no. 13, pp. 3715–3726, 2008, doi: 10.1080/01431160701772542.

[12] W. Li, J. D. M. Saphores, and T. W. Gillespie, "A comparison of the economic benefits of urban green spaces estimated with NDVI and with high-resolution land cover data," *Landsc. Urban Plan.*, vol. 133, pp. 105–117, 2015, doi: 10.1016/j.landurbplan.2014.09.013.

[13] M. Hu and B. Xia, "A significant increase in the normalized difference vegetation index during the rapid economic development in the Pearl River Delta of China," *L. Degrad. Dev.*, vol. 30, no. 4, pp. 359–370, 2019, doi: 10.1002/ldr.3221.

[14] Y. Guo *et al.*, "Spatial and temporal changes in vegetation in the Ruoergai Region, China," *Forests*, vol. 12, no. 1, pp. 1–17, 2021, doi: 10.3390/f12010076.

[15] G. Wang and W. Peng, "Detecting influences of factors on GDP density differentiation of rural poverty changes," *Struct. Chang. Econ. Dyn.*, vol. 56, pp. 141–151, 2021, doi: 10.1016/j.strueco.2020.10.004.

[16] Z. M. Fan, X. Y. Bai, and N. Zhao, "Explicating the responses of NDVI and GDP to the poverty alleviation policy in poverty areas of China in the 21st century," *PLoS One*, vol. 17, no. 8 August, pp. 1–16, 2022, doi: 10.1371/journal.pone.0271983.

[17] H. W. Duolaiti, X., Kasim, A., Reheman, R., & Laing, "Water body extraction of Ebinur Lake based on four water indexes and analysis of spatial-temporal changes," *J. Yangtze River Sci. Res. Inst.*, 2022.

[18] S. Ahmed, "Assessment of urban heat islands and impact of climate change on socioeconomic over Suez Governorate using remote sensing and GIS techniques," *Egypt. J. Remote Sens. Sp. Sci.*, vol. 21, no. 1, pp. 15–25, 2018, doi: 10.1016/j.ejrs.2017.08.001.

[19] A. D. Sakti *et al.*, "School location analysis by integrating the accessibility, natural and biological hazards to support equal access to education," *ISPRS Int. J. Geo-Information*, vol. 11, no. 1, 2022, doi: 10.3390/ijgi11010012.

[20] X. Zhang *et al.*, "Long-term trends in NO2 columns related to economic developments and air quality policies from 1997 to 2016 in China," *Sci. Total Environ.*, vol. 639, no. 2, pp. 146–155, 2018, doi: 10.1016/j.scitotenv.2018.04.435.

[21] A. Sinha, "Trilateral association between SO2/NO2 emission, inequality in energy intensity, and economic growth: A case of Indian cities," *Atmos. Pollut. Res.*, vol. 7, no. 4, pp. 647–658, 2016, doi: 10.1016/j.apr.2016.02.010.

[22] D. Dissanayake, T. Morimoto, Y. Murayama, M. Ranagalage, and H. H. Handayani, "Impact of urban surface characteristics and socio-economic variables on the spatial variation of land surface temperature in Lagos City, Nigeria," *Sustain.*, vol. 11, no. 1, pp. 1–23, 2019, doi: 10.3390/su11010025.

[23] M. Khal, A. Algouti, A. Algouti, N. Akdim, S. A. Stankevich, and M. Menenti, "Evaluation of open digital elevation models: estimation of topographic indices relevant to erosion risk in the Wadi M'Goun watershed, Morocco," *AIMS Geosci.*, vol. 6, no. 2, pp. 231–257, 2020, doi: 10.3934/geosci.2020014.

[24] Q. Chen *et al.*, "Mapping China ' s regional economic activity by integrating points-of-interest and remote sensing data with random forest," vol. 0, no. 0, pp. 1–19, 2020, doi: 10.1177/2399808320951580.

[25] H. Liu *et al.*, "Erratum to: Nightlight as a proxy of economic indicators: Fine-grained GDP inference around Chinese mainland via attention-augmented CNN from daytime satellite imagery (Remote SENS. 2021, 13, 2067)," *Remote Sens.*, vol. 13, no. 17, 2021, doi: 10.3390/rs13173360.

[26] Z. Chen, W. Wang, H. Zong, and X. Yu, "Combining the NPP-VIIRS-like dataset and sentinel-2 images," 2024.

[27] N. Wu, J. Yan, D. Liang, Z. Sun, R. Ranjan, and J. Li, "High-resolution mapping of GDP using multi-scale feature fusion by integrating remote sensing and POI data," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 129, no. April, p. 103812, 2024, doi: 10.1016/j.jag.2024.103812.

[28] Badan Pusat Statistik Provinsi Jawa Timur, "Produk domestik regional bruto Provinsi Jawa Timur menurut lapangan usaha 2018-2022," *Badan Pusat Statistik Jawa Timur*, 2023. .

[29] A. De Mauro, M. Greco, and M. Grimaldi, "What is big data? A consensual definition and a review of key research topics," *AIP Conf. Proc.*, vol. 1644, pp. 97–104, 2015, doi: 10.1063/1.4907823.
[30] J. G. Lee and M. Kang, "Geospatial big data: Challenges and opportunities," *Big Data Res.*, vol. 2, no. 2, pp. 74–81, 2015, doi: 10.1016/j.bdr.2015.01.003.
[31] BPS RI, *Teknik pengumpulan data dan preprocessing citra satelit.* Badan Pusat Statistik, 2022.
[32] R. A. Schowengerdt, *Remote sensing.* Elsevier, 2007.
[33] J. B. Campbell, *Introduction to remote sensing*. Guilford Press, 2011.
[34] G. K. Moore, "What is a picture worth? a history of remote sensing," *Hydrol. Sci. Bull.*, vol. 24, no. 4, pp. 477–485, 1979, doi: 10.1080/02626667909491887.
[35] N. Suwargana, "Temporal dan spektral pada citra satelit landsat, spot Dan ikonos," *J. Ilm. Widya*, vol. 1, no. 2, pp. 167–174, 2013.
[36] A. F. Syah, "Penginderaan jauh dan aplikasinya di wilayah pesisir dan lautan," *J. Kelaut.*, vol. 3, no. 1, pp. 18–28, 2010.
[37] A. Oktaviani and Y. Johan, "Perbandingan resolusi spasial, temporal dan radiometrik serta kendalanya," *J. Enggano*, vol. 1, no. 2, pp. 74–79, 2016, doi: 10.31186/jenggano.1.2.74-79.
[38] C. D. Elvidge, M. Zhizhin, F.-C. Hsu, and K. Baugh, "What is so great about nighttime VIIRS data for the detection and characterization of combustion sources?," in *Proceedings of the Asia-Pacific Advanced Network*, 2013, vol. 35, no. 0, p. 33, doi: 10.7125/apan.35.5.
[39] X. Chen and W. D. Nordhaus, "VIIRS nighttime lights in the estimation of cross-sectional and time-series GDP," *Remote Sens.*, vol. 11, no. 9, pp. 1–11, 2019, doi: 10.3390/rs11091057.
[40] R. Ramadhan, A. W. Wijayanto, and S. Pramana, "Geospatial big data approaches to estimate granular level poverty distribution in East Java, Indonesia using machine learning and deep learning regressions," in *Proceedings of The International Conference on Data Science and Official Statistics, 2023(1)*, 2023, pp. 186–200, doi: https://doi.org/10.34123/icdsos.v2023i1.359.
[41] A. W. Wijayanto, D. W. Triscowati, and A. H. Marsuhandi, "Maize field area detection in East Java, Indonesia: An integrated multispectral remote sensing and machine learning approach," *ICITEE 2020 - Proc. 12th Int. Conf. Inf. Technol. Electr. Eng.*, pp. 168–173, 2020, doi: 10.1109/ICITEE49829.2020.9271683.
[42] Y. Li, Z. Cao, H. Long, Y. Liu, and W. Li, "Dynamic analysis of ecological environment combined with land cover and NDVI changes and implications for sustainable urban–rural development: The case of Mu Us Sandy Land, China," *J. Clean. Prod.*, vol. 142, pp. 697–715, Jan. 2017, doi: 10.1016/j.jclepro.2016.09.011.
[43] Bo-Cai Gao, "NDWI A normalized difference water index for remote sensing of vegetation liquid water from space," *Elsevier*, vol. 266, no. April, Feb. 1996.
[44] S. K. McFeeters, "The use of the normalized difference water index (NDWI) in the delineation of open water features," *Int. J. Remote Sens.*, vol. 17, no. 7, pp. 1425–1432, 1996, doi: 10.1080/01431169608948714.
[45] H. Krishna Karanam, "Study of normalized difference built-up (Ndbi) index in automatically mapping urban areas from landsat TM imagery," *Int. J. Sci. Res. Rev.*, vol. 7, no. 1, pp. 1–8, 2018.
[46] Y. Zha, J. Gao, and S. Ni, "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery," *Int. J. Remote Sens.*, vol. 24, no. 3, pp. 583–594, 2003, doi: 10.1080/01431160304987.
[47] C. He, P. Shi, D. Xie, and Y. Zhao, "Improving the normalized difference built-up index to map urban built-up areas using a semiautomatic segmentation approach," *Remote Sens. Lett.*, vol. 1, no. 4, pp. 213–221, 2010, doi: 10.1080/01431161.2010.481681.
[48] E. D. Meutia, M. Fikri, R. Munadi, and Yunida, "Analisis data citra satelit terra modis sebagai indikasi potensi ikan di wilayah kreung raya," *KITEKRO J. Komputer, Teknol. Informasi, dan Elektro*, vol. 7, no. 1, pp. 9–14, 2022.
[49] J. P. Veefkind *et al.*, "TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications," *Remote Sens. Environ.*, vol. 120, no. 2012, pp. 70–83, 2012, doi: 10.1016/j.rse.2011.09.027.
[50] R. D. Stephens and S. H. Cadle, "Remote sensing measurements of carbon monoxide emissions from on-road vehicles," *J. Air Waste Manag. Assoc.*, vol. 41, no. 1, pp. 39–46, 1991, doi: 10.1080/10473289.1991.10466823.
[51] C. Nistor, M. Vîrghileanu, and B. Mihai, "Monitoreo de la contaminación por dióxido de nitrógeno (NO2) con imágenes satelitales Sentinel-5P en Europa durante el brote pandémico de

coronavirus," no. 2, 2020.
[52] R. Kurniawan, R. Hasabi, S. K. Wongsonadi, P. U. Gio, A. Purwanto, and B. Sumargo, "Examining the influence of congestion, industry, and green open space on air quality vulnerability in towards green development in Jakarta," *Innov. Green Dev.*, vol. 4, no. 3, 2025, doi: 10.1016/j.igd.2025.100247.
[53] K. Bakhsh, T. Akmal, T. Ahmad, and Q. Abbas, "Investigating the nexus among sulfur dioxide emission, energy consumption, and economic growth: empirical evidence from Pakistan," *Environ. Sci. Pollut. Res.*, vol. 29, no. 5, pp. 7214–7224, 2022, doi: 10.1007/s11356-021-15898-9.
[54] L. Shikwambana, P. Mhangara, and M. Kganyago, "Assessing the relationship between economic growth and emissions levels in south africa between 1994 and 2019," *Sustain.*, vol. 13, no. 5, pp. 1–15, 2021, doi: 10.3390/su13052645.
[55] M. Abrams, R. Crippen, and H. Fujisada, "ASTER global digital elevation model (GDEM) and ASTER global water body dataset (ASTWBD)," *Remote Sens.*, vol. 12, no. 7, pp. 1–12, 2020, doi: 10.3390/rs12071156.
[56] I. Elkhrachy, "Vertical accuracy assessment for SRTM and ASTER digital elevation models: A case study of Najran city, Saudi Arabia," *Ain Shams Eng. J.*, vol. 9, no. 4, pp. 1807–1817, 2018, doi: 10.1016/j.asej.2017.01.007.
[57] L. W. Yeow, R. Low, Y. X. Tan, and L. Cheah, "Point-of-interest (Poi) data validation methods: An urban case study," *ISPRS Int. J. Geo-Information*, vol. 10, no. 11, 2021, doi: 10.3390/ijgi10110735.
[58] A. P. Utomo *et al.*, "Regresi robust untuk memodelkan pendapatan usaha industri makanan non-makloon berskala mikro dan kecil di Jawa Barat tahun 2013," pp. 63–74, 2013.
[59] Soemartini, *Pencilan (Outliers)*. Jatinangor: Universitas Padjajaran, 2007.
[60] J. Yohai, D. B. Aires, and C. E. Ma, "High breakdown-point and high efficiency robust estimates for regression," vol. 15, no. 2, pp. 642–656, 2014.
[61] J. Raymaekers and P. J. Rousseeuw, "Transforming variables to central normality: Machine learning," *Mach. Learn.*, vol. 113, no. 8, pp. 4953–4975, 2024, doi: 10.1007/s10994-021-05960-5.
[62] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. 2011.
[63] W. Lin, "Temporal and spatial analysis of integrated energy and environment efficiency in China based on a green GDP index," pp. 1376–1390, 2011, doi: 10.3390/en4091376.
[64] Y. Wang and Y. Chen, "Using VPI to measure poverty-stricken villages in China," *Soc. Indic. Res.*, no. June, 2016, doi: 10.1007/s11205-016-1391-5.
[65] S. R. Putri and A. W. Wijayanto, "Developing relative spatial poverty index using integrated remote sensing and geospatial big data approach : A case study of East Java , Indonesia," 2024.
[66] R. Ramadhan and A. W. Wijayanto, "Integrating satellite imageries and multiple geospatial big data for granular mapping of spatial distribution of the human development index in East Java , Indonesia," in *Proceedings of The International Conference on Data Science and Official Statistics*, 2023, vol. 2023(1), pp. 274–295, doi: https://doi.org/10.34123/icdsos.v2023i1.369.
[67] F. Kartiasih, N. D. Nachrowi, I. D. Gede, and K. Wisana, "Information technology for development inequalities of Indonesia ' s regional digital development and its association with socioeconomic characteristics : a spatial and multivariate analysis," *Inf. Technol. Dev.*, vol. 0, no. 0, pp. 1–30, 2022, doi: 10.1080/02681102.2022.2110556.
[68] A. Cartone and P. Postiglione, "Principal component analysis for geographical data : the role of spatial effects in the definition of composite indicators Principal component analysis for geographical data : the role of spatial effects in the de fi nition of composite," *Spat. Econ. Anal.*, vol. 0, no. 0, pp. 1–22, 2020, doi: 10.1080/17421772.2020.1775876.
[69] Y. Nurmasari and A. W. Wijayanto, "Oil palm plantation detection in Indonesia using sentinel-2 and landsat-8 optical satellite imagery (Case Study: Rokan Hulu Regency, Riau Province)," *Int. J. Remote Sens. Earth Sci.*, vol. 18, no. 1, p. 1, 2021, doi: 10.30536/j.ijreses.2021.v18.a3537.
[70] D. Coondoo, I. Researcher, and A. Majumder, "District-level poverty estimation : A proposed method District-level poverty estimation : a proposed method," no. October 2014, 2011, doi: 10.1080/02664763.2010.547568.
[71] L. Mládková, V. Pení, and R. Va, "Forest soil acidification assessment using principal component analysis and geostatistics," vol. 140, pp. 374–382, 2007, doi: 10.1016/j.geoderma.2007.04.018.
[72] F. J. Riggins, "The digital divide : Current and future research directions," vol. 6, no. 12, 2005, doi: 10.17705/1jais.00074.