Feasibility study of Satellite Imagery Analysis for Wealth Index Development in Indonesia

Organisation: Statistics Indonesia (BPS)
Author(s): Arie Wahyu Wijayanto
Date: 20-10-2021
Version: 1.0

1. Background

The wealth index is a composite measure of a household's cumulative living standard. The existing wealth index has been conventionally compiled using data from field surveys calculated using easy-to-collect data on a household's ownership of selected assets, such as televisions and bicycles; materials used for housing construction; and types of water access and sanitation facilities. In Indonesia, the wealth index is used for stratification purposes. It will increase the efficiency of surveys and get more representative samples. In this project, we are working to examines an alternative method in estimating those indexes using readily available geospatial data with remote sensing approaches. We investigate the benefit of machine learning (ML) to produce the proximity of wealth index from night-time light intensity of satellite imageries. The task of machine learning models is to classify the selected area into corresponding classes in wealth index area level. The official statistical surveys published by BPS-Statistics Indonesia are used as the ground truth labels to validate our proposed model. Hopefully, the best found classifier and features will be promising to provide the essential indicators for the statistical surveys and census in poverty and wealth mapping.

2. Data

2.1 Input data

We investigate the feasibility of using publicly accessible satellite images to enhance the granularity of poverty statistics compiled using conventional estimation methods. Many existing literatures discusses the advantage of daytime images and nighttime images of a particular region as its geospatial features and uses the official poverty statistics as input data. Hence, we use the following data in this study.

- Daytime Satellite Images

We use publicly available satellite images in this study, specifically optical medium resolution satellites, i.e. Landsat-8 (15-meter resolution) and Sentinel-2 (10-meter resolution). Those satellites offers georeferenced features with high frequency of updates compared to the expensive conventional field survey.

- Nighttime Satellite Images

The required data consist of the National Polar-orbiting Partnership - Visible Infrared Imaging Radiometer Suite (NPP-VIIRS) satellite images which captures the night-time light observation. NPP-VIIRS data are publicly accessible.

- Official Poverty Database

The integrated national-scale Indonesia poverty database (PBDT) 2015 is used, as well as other relevant newer official poverty estimates in 2018 and 2020

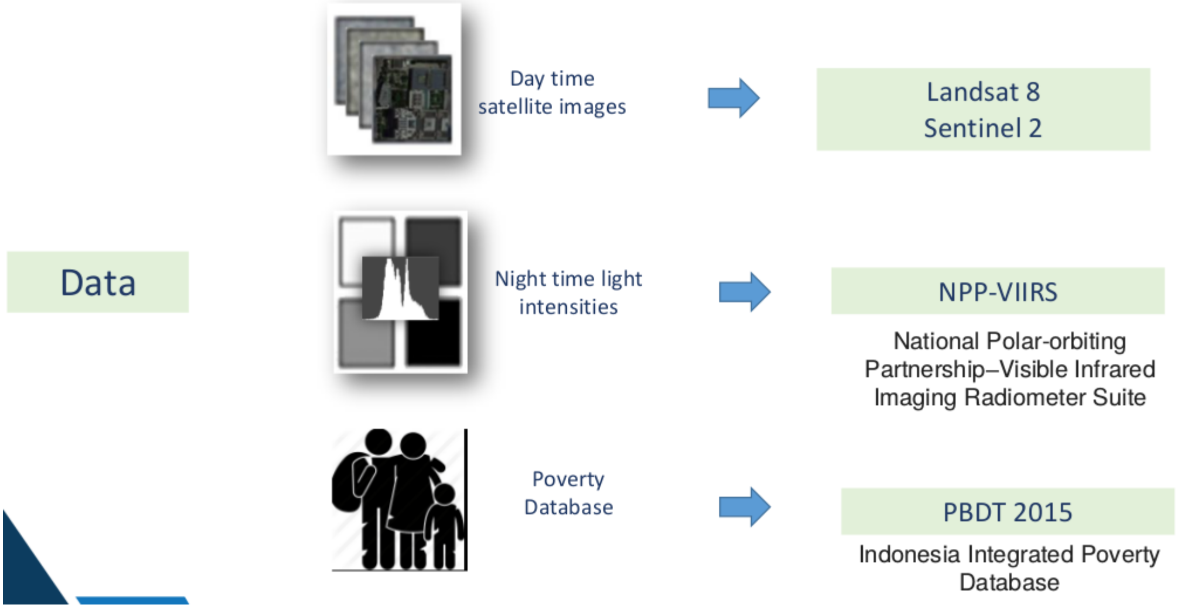


Figure 1. Data

2.2 Data preparation

Data preparation include the data cleaning process on satellite imageries, such as cloud cover removal, consitency checking, etc. Each satellite images are then sliced into grid structures with size of 1km or 4km depending on the area of observation. Larger province area are sliced in 4km grids, while smaller provinces use 1km grids.

2.3 Feature selection

Multiple spectral bands and indices are evaluated and selected. In Landsat-8 and Sentinel-2, not only the visible spectral bands that will be evaluated, but also other bands such as near infrared, panchromatic bands, etc.

2.4 Output data

Output data are shapefiles of sliced satellite images of three sources (VIIRS, Landsat-8, and Sentinel-2) which accompanied by a georeferenced table of its features, center coordinates (longitudes and latitudes), and poverty data labels taken from official poverty data.

3. Machine learning solution

3.1 Models tried

Geospatial data from satellite imageries are naturally unstructured, noisy, and expensive to process both statistically and computationally. Nowadays, deep learning, the Convolutional Neural Networks (CNNs) architecture have gained state-of-the-art performances in the field of computer vision. This offered opportunities to utilize deep learning on satellite images to gain information from the ground. Structurally, a neural network is constructed from numerous nodes and edges. A node can be a variable or a mathematical function connected by edges. These nodes combine together to form different layers within the neural network. The input layer takes in the raw input data. In the hidden layers, each node or neuron serves as filter and is activated each time it detects a specific pattern or feature. The output layer simply organizes the identified features into an appropriate category. We use a Convolutional neural networks architecture, i.e. the Resnet-50 in our study.

3.2 Model(s) finally selected and quality criteria used

The deep learning architecture, Convolutional Neural Networks are being our consideration. Among other, the Resnet-50 architecture by Fast.AI is then selected. The quality measurement to compare different ML models include the accuracy and F-1 score. The best performed models on those measurements will be finally selected.

3.3 Hardware used

The cloud service provided by Google Earth Engine and Google Collab are extensively used with the additional use of Python libraries. Office hardware of consumer grade PCs and laptops, at BPS-Statistics Indonesia are heavily involved during the process.

4. Results

Figure 2 illustrates the methodological framework of our investigation.

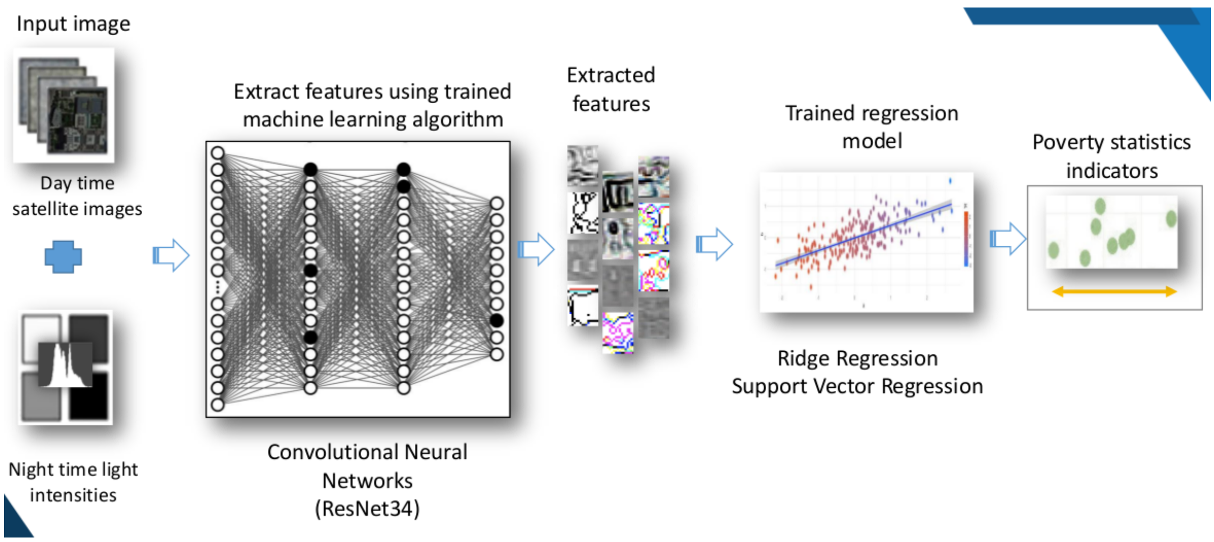


Figure 2. Methodological framework

For the sake of brevity and simplicity, in this report, we show the Provinces of DI Yogyakarta and Banten as case study to further examine our prediction model.

- Case Study on DI Yogyakarta Province

Night-Time Lights Luminosity

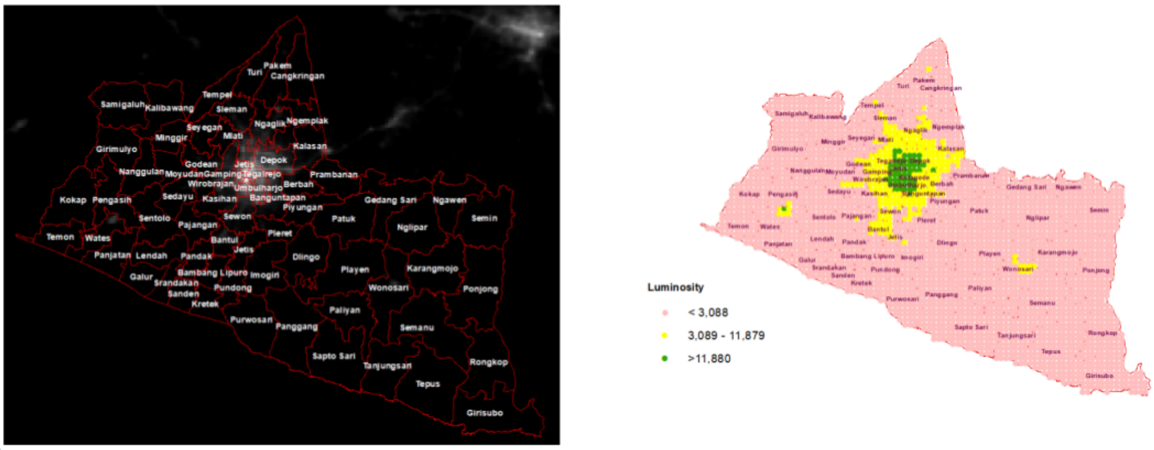


Figure 3. The center of Yogyakarta City has a greater luminosity intensity than rural areas and areas outside the city

Night-Time Lights Luminosity **Official Poverty Distribution (PBDT 2015)**

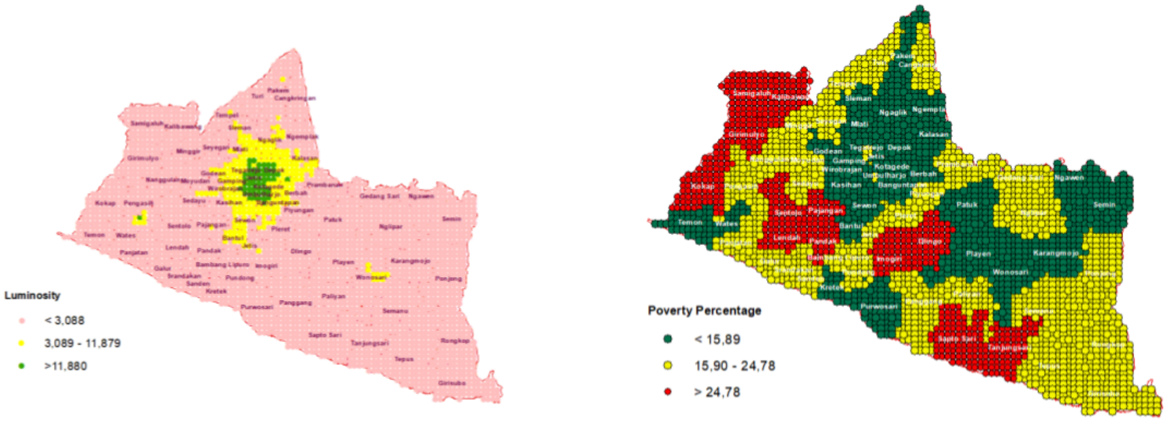
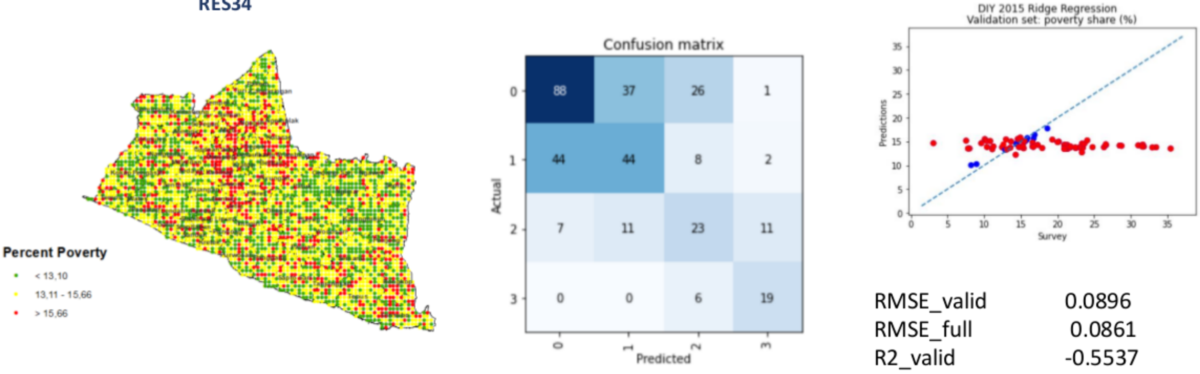


Figure 4. The capital of DI Yogyakarta province, namely Yogyakarta city, has a greater intensity of night-time light and a lower poverty rate than other areas

A. Predicted Poverty Distribution (Combination of Day Time & Night Time Light)

Poverty Percentage by prediction model with RES34 **CNN Model Testing and Evaluation**



RMSE_valid	0.0896
RMSE_full	0.0861
R2_valid	-0.5537
R2_full	-0.4796

Figure 5. The resulted model predictions when compared with the Official Poverty Distribution (PBDT 2015)

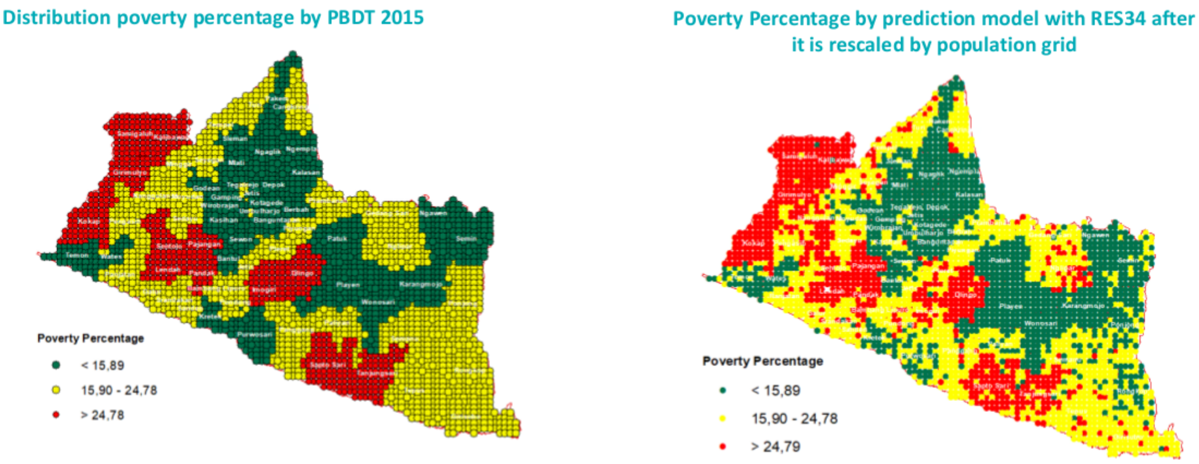


Figure 6. The results of the model predictions after rescaling are quite good in estimating regional poverty with an RMSE value of 8 percent

B. Predicted Poverty Distribution (Night Time Light)

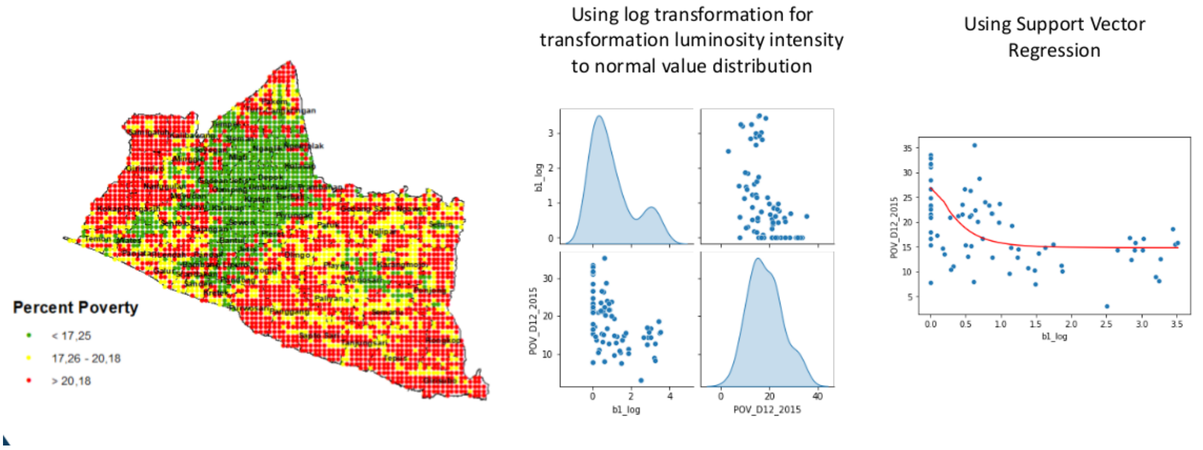


Figure 7. The results of the model predictions with only night-time light imagery are better than the combination of both day-time and night-time light imagery.

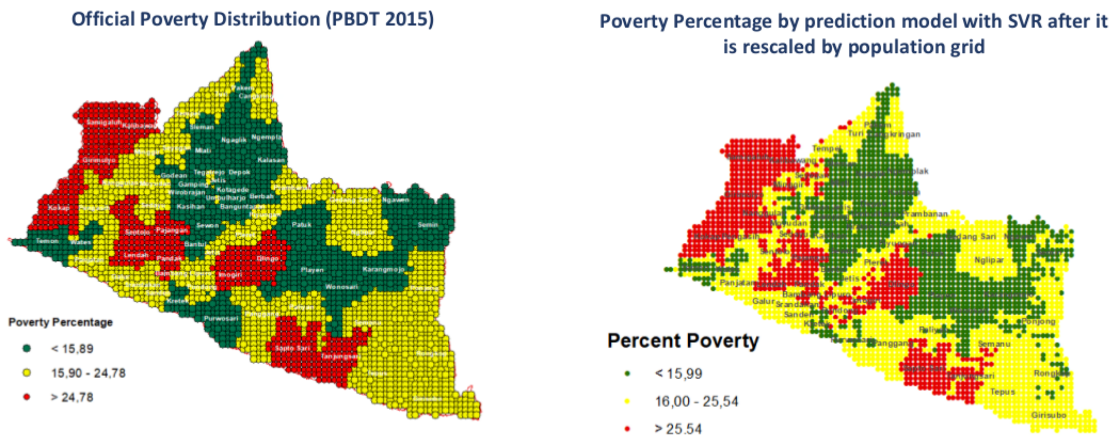
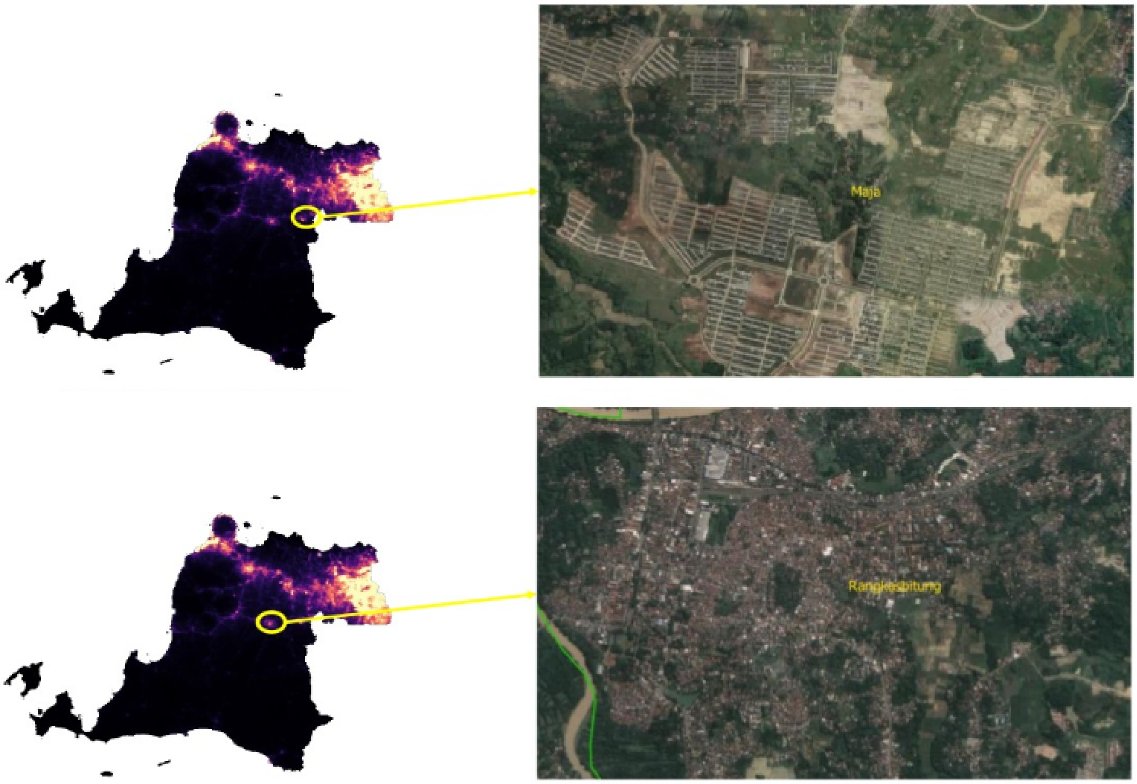


Figure 8. Using NTL Imagery, we can predict regional poverty distribution more properly. After rescaling using the population grid, we better spatial distribution poverty compared with the Official Poverty Database (PBDT 2015).

- Case Study on Banten Province

We conduct a similar investigation and build the same prediction model to Banten province. The following figure illustrates how we use the night-time satellite images as an input to estimate the poverty distribution.

As we identified some locations require further investigations. Based on observations made through night satellite imagery, it appears that there are several points where they appear as the brightest positions, but dark for the surrounding area. From this anomaly, it is necessary to conduct field checks to photograph the phenomena captured by the night image. Based on previous observations, using satellite imagery, several points have this anomaly.



The following coordinates are then selected to be checked with the ground-checking.

ADM2_EN	ADM2_PCODE	latitude	longitude	rwi	ntl	prediksi_no_cnn	prediksi_cnn
Lebak	ID3602	-6.3480564	106.3806152	0.354	12.733365	0.422642739	0.520328063
Lebak	ID3602	-6.3480564	106.4025879	0.453	3.9811645	0.471481456	0.139186679
Lebak	ID3602	-6.3480564	106.3586426	0.276	4.638716	0.387174166	0.354541393
Lebak	ID3602	-6.3480564	106.3806152	0.354	12.733365	0.422642739	0.520328063
Lebak	ID3602	-6.3480564	106.4025879	0.453	3.9811645	0.471481456	0.139186679
Lebak	ID3602	-6.3480564	106.3586426	0.276	4.638716	0.387174166	0.354541393

From these locations, some locations are most likely to be visited because they relate to the time and distance required. The following examples of two locations are the residential cluster area and the area around the night market.

The ground-checking results indicates that the satellite data and prediction models are good and satisfactory.

Key Findings

• The model has been able to estimates the poverty spatial distribution relatively well.

• Ground checks have been conducted to ensure that the night light data represents the economic activities in some area. For instance in Banten Province.