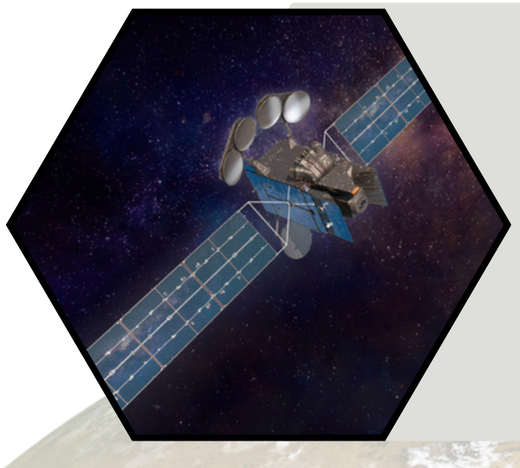
DOSM/NBDAC/1.2025/SERIES 91

MINISTRY OF ECONOMY
DEPARTMENT OF STATISTICS MALAYSIA
NEWSLETTER

EARTH OBSERVATION:
ASSESSING THE POTENTIAL OF SATELLITE IMAGERY FOR POVERTY ANALYSIS

*This project is in the R&D stage and currently focuses on the Kelantan region.

SATELLITE IMAGE



Satellite images are pictures of the Earth taken from space that show **land, water, vegetation, cities,** and **human activities,** helping us to understand environmental and socio-economic conditions.

POVERTY

Poverty is a condition in which individuals or communities **lack sufficient resources** such as income, education, health, housing, and access to services to maintain an acceptable standard of living.

WORKFLOW METHODOLOGY

1. **Data Acquisition**
Gathered satellite, POI, and census data.

2. **Pre-Processing Data**
Cleaned, merged, and filled missing values.

3. **Data Transformation**
Standardized data for modelling.

4. **Variable Selection**
Choose key variables using statistical methods.

5. **Modelling**
Built models to predict poverty levels.

6. **Performance Evaluation**
Tested model accuracy using performance metrics.

SOURCE OF DATA

SATELLITE IMAGES DATA

Using Google Earth Engine, 9 types of satellite images were collected:

- Showing light at night (**NTL**)
- Green cover (**NDVI**)
- Water (**NDWI**)
- Built-up areas (**BUI**)
- Land temperature
  - day (**LST-DAY**)
  - night (**LST-NIGHT**)
- Air pollution
  - Carbon Monoxide (**CO**)
  - Nitrogen Dioxide (**NO2**)
  - Sulfur Dioxide (**SO2**)

POINT OF INTEREST(POI) DATA

Points of Interest (POI) were extracted from **OpenStreetMap** to understand the types of places in each area. These include **amenities** (like schools, shops, restaurants), **healthcare** (clinics, hospitals), **public transport** (bus stops, stations), **buildings**, **industrial** and **land use areas**, and etc. POI data helps show how accessible, developed, and active a community is.

CENSUS DATA

Uses Census Data:

- Provides detailed **social and economic information** for each area at granular level.
- Includes population, area size, housing, employment, household income and spending, education, health access, basic amenities, business establishments, and public facilities.

WHY WE CHOOSE THIS DATA?

These data sources are **affordable** and can be **updated often**, unlike traditional surveys that are expensive and slow (Putri, Wijayanto, & Pramana, 2023). However, using only satellite and POI data can make the model less reliable. To **improve accuracy**, census data is added because it provides real information about people's lives, such as population, income, housing, and access to facilities.

SATELLITE IMAGES DATA

NIGHTTIME LIGHT (NTL)

Nighttime light imagery from sensors such as DMSP-OLS and VIIRS has become one of the most robust Earth Observation proxies for economic development and poverty mapping.

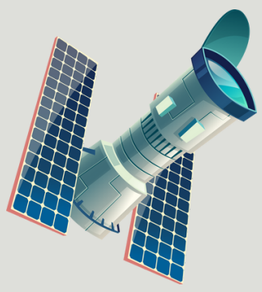
NORMALIZED DIFFERENCE VEGETATION INDEX (NDVI)

Vegetation indices such as the **Normalized Difference Vegetation Index (NDVI)** and Enhanced Vegetation Index (EVI) quantify photosynthetic activity and land productivity.

NORMALIZED DIFFERENCE WATER INDEX (NDWI)

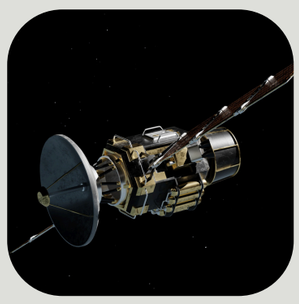
The Normalized Difference Water Index (NDWI) and Modified NDWI (MNDWI) are important for identifying areas prone to flooding or drought.

BUILT-UP INDEX (BUI)



Built-up indices such as the Normalized Difference Built-Up Index (NDBI) and Built-Up Index (BUI) **quantify urban expansion and infrastructure density**.

LAND SURFACE TEMPERATURE (LST)



Land Surface Temperature derived from MODIS or Landsat data provides insight into environmental **comfort**, **energy use**, **and urban heat-island effects.**

AIR POLLUTION

Recent advances in Sentinel-5P data allow monitoring of **Carbon Monoxide (CO)**, **Nitrogen Dioxide (NO_2)**, and **Sulfur Dioxide (SO_2)**. These pollutants are indirect proxies for industrial activity, traffic density, and household energy usage factors intertwined with socioeconomic conditions.

PRE-PROCESSING DATA

Before building the model to predict poverty, all the collected data needed to be **cleaned** and **prepared**. This step ensures the information is accurate and consistent. Some data had **missing values**, which means certain areas didn't have complete information. To fix this, the missing numbers were filled using the **average value** of similar regions, so that no data point was left empty. After that, all datasets which is satellite image, POI, and census data, were **combined into one** complete dataset, making it easier for the model to learn and make predictions.

DATA TRANSFORMATION

NORMALITY TEST

- Used to check if the data follows a normal (**balanced**) pattern.
- A normal pattern means most values are **close to the average**, with few **extreme values**.
- Important because some models work better with normally distributed data.
- Based on the obtained result, it can demonstrated that the nature of the data itself is **not normally distributed**.

DATA TRANSFORMATION

- Transformed the data using the **Yeo–Johnson** transformation to normalize the distribution and ensure comparability among variables.
- Applied the Yeo–Johnson transformation to **reduce skewness** and **stabilize variance** before modelling.
- Standardized and transformed the data using the Yeo–Johnson method to **improve model accuracy and consistency**.

WHY WE DO THIS?

Before building the model, the data needed to be **aligned** so that all variables could be compared fairly. Some data values were **uneven**, for example, certain numbers were much higher or lower than others. To fix this, we used the **Yeo–Johnson transformation**, which helps make the data **more balanced** and **stable**. This step makes it easier for the model to learn patterns accurately and produce more reliable results.

TOOLS AND METHOD

- **Google Earth Engine (GEE)**: Used to collect and process satellite images.
- **QGIS**: Used to visualize and analyze geographic data on maps.
- **Python**: Used to clean the data, select important variables, and build the machine learning model.
- **Correlation Analysis**: Used to identify relationships between different factors that influence poverty.

REFERENCE

- Putri, S. R., Wijayanto, A. W., & Pramana, S. (2023). Multi-source satellite imagery and point of interest data for poverty mapping in East Java, Indonesia: Machine learning and deep learning approaches.