Remote Sensing Combined with Machine Learning for Poverty Estimation and Mapping-Zambia

Owen Siyoto Supervised by Dr. Meltenisova Ekaterina

Novosibirsk State University

o.siyoto@g.nsu.ru

Master Thesis

December 26, 2019

Overview

Goal and Objectives

- GOAL
  - The Goal of the research is to estimate the levels of poverty in Zambia with the help machine learning and remote sensing.
- Objectives
  - To use machine learning methods for economic analysis
  - To use remote sensing data in poverty estimation
  - To estimate poverty for Zambia at lower administrative levels

Object and Subject

- OBJECT
  - Country Zambia
- Subject
  - We considered the level of poverty in Zambia;

Motivation

Goal number 1 of the sdgs is “end poverty in all its forms everywhere”.

It is estimated that One Billion people, or 11 percent of the world population, still live in extreme poverty (Less than 2 Dollar per day).

The lack of reliable data in developing countries is a major obstacle to sustainable development, food security, and disaster relief.

Similar Works

- Neal Jean, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, Stefano Ermon; S 2016. Combining satellite imagery and machine learning to predict poverty
- Xie, M.; Jean, N.; Burke, M.; Lobell, D.; and Ermon, S. 2015. Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping
- Xizhi Zhao, Bailang Yu , Yan Liu , Zuoqi Chen, Qiaoxuan Li, Congxiao Wang and Jianping Wu; S 2019. Estimation of Poverty Using Random Forest Regression with Multi-Source Data: A Case Study in Bangladesh

World and Africa Night Lights

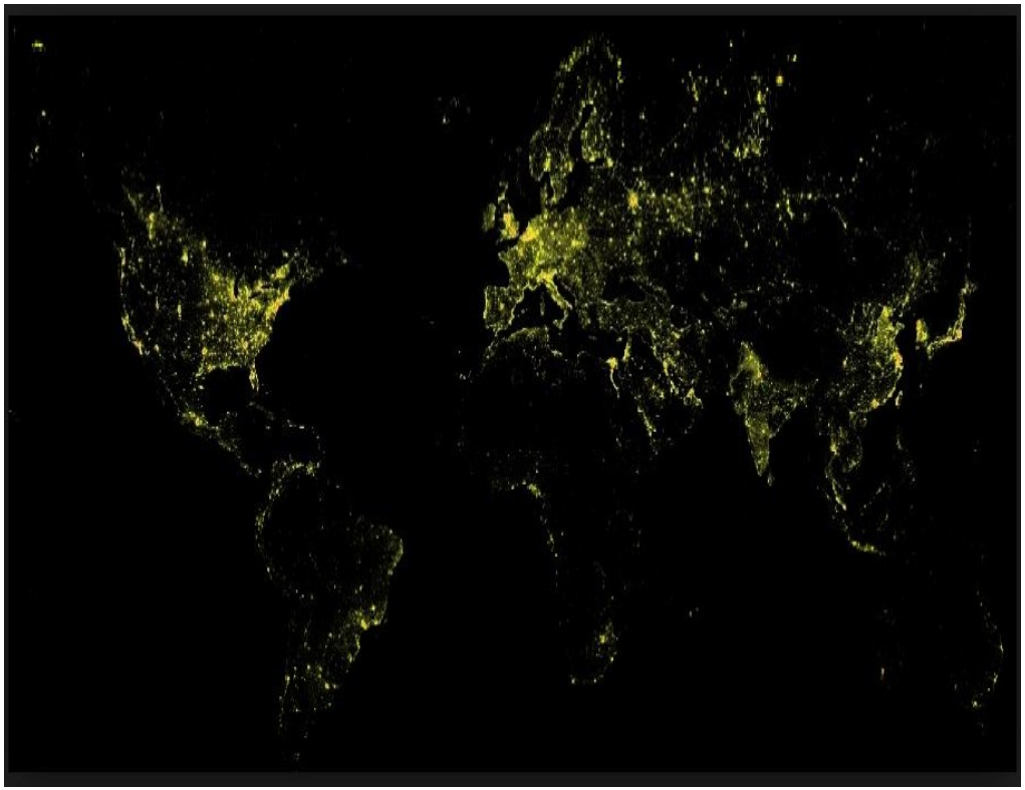

Figure: World Night Lights

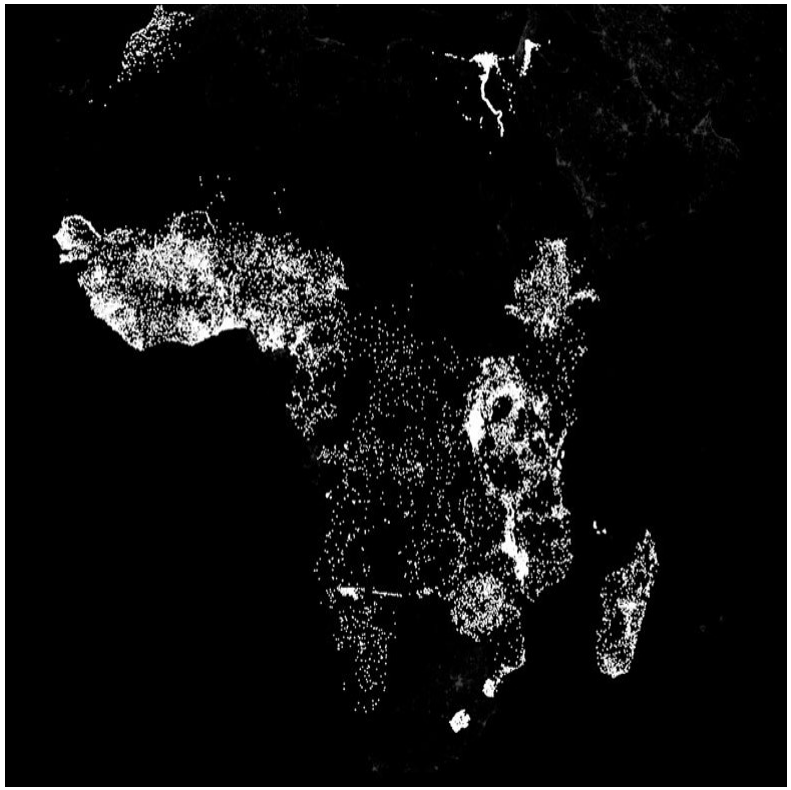

Figure: Africa Night Lights

Methodology

- We shall construct a linear chain transfer learning graph with : V = (P1, P2, P3) and E = (P1, P2), (P2, P3).
- The first transfer learning problem P1 is object recognition on ImageNet
- The second problem P2 is predicting nighttime light intensity from daytime satellite imagery, simultaneously learning features that are useful for poverty prediction;
- The third problem P3 is predicting poverty from daytime satellite imagery.

Proposed Methodology Simplified Pipeline

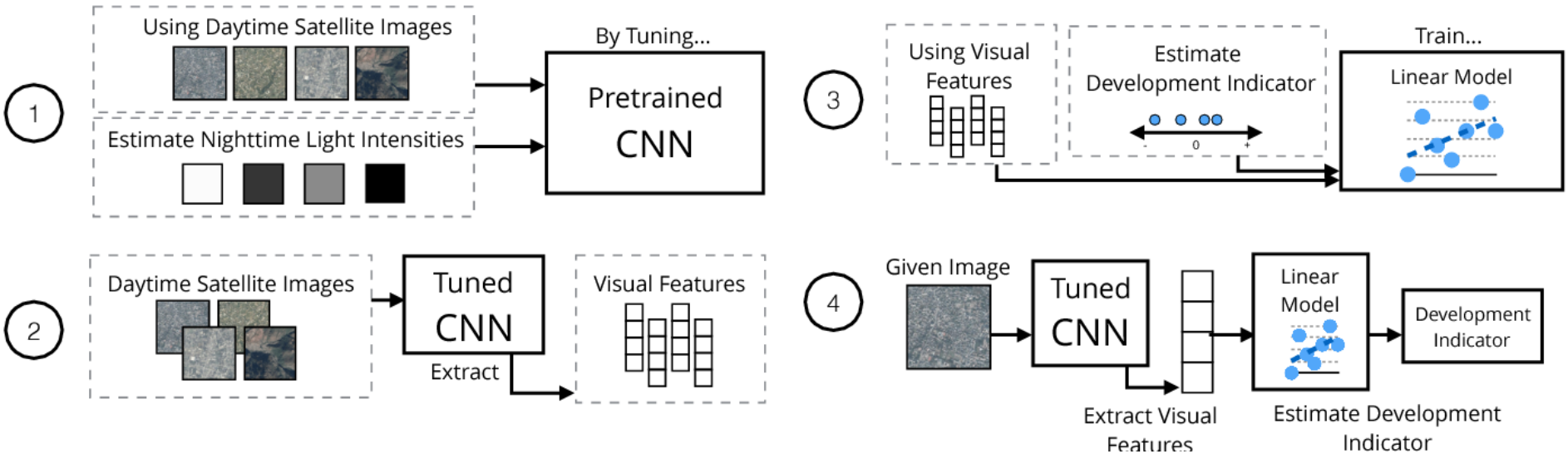

Figure: Pipeline.

Methods Explained

- First Step
  - First, we start with a convolutional neural network (CNN) model that has been pretrained on ImageNet. In learning to classify each image correctly, the model learns to identify low level image features such as edges and corners.
- Second Step
  - Next, we build on the knowledge gained from this image classification task and fine-tune the CNN on a new task, training it to predict the nighttime light intensities corresponding to input daytime satellite imagery. The trained CNN can be treated as a feature extractor that has learned a nonlinear mapping from each input image to a concise feature vector representation.
- Third Step
  - In the final step, we use mean cluster-level values from the survey data along with the corresponding image features extracted from daytime imagery by the CNN to train ridge regression models that can estimate cluster-level expenditures/Poverty levels. Regularization in the ridge model guards against overfitting, a potential challenge given the high dimensionality of the extracted features and the relatively small survey

Evaluation Metrics

- R-squared
  - It is a statistical measure that represents the proportion of the variance for a target variable (Poverty rate) that is explained by the independent variables (Extracted features). R-squared values range from 0 to 1 and are commonly stated as percentages.
- Root Mean Squared Error (RMSE)
  - The average magnitude of the residuals or error.

Step 2 - Visualization of features(Further Illustrated)

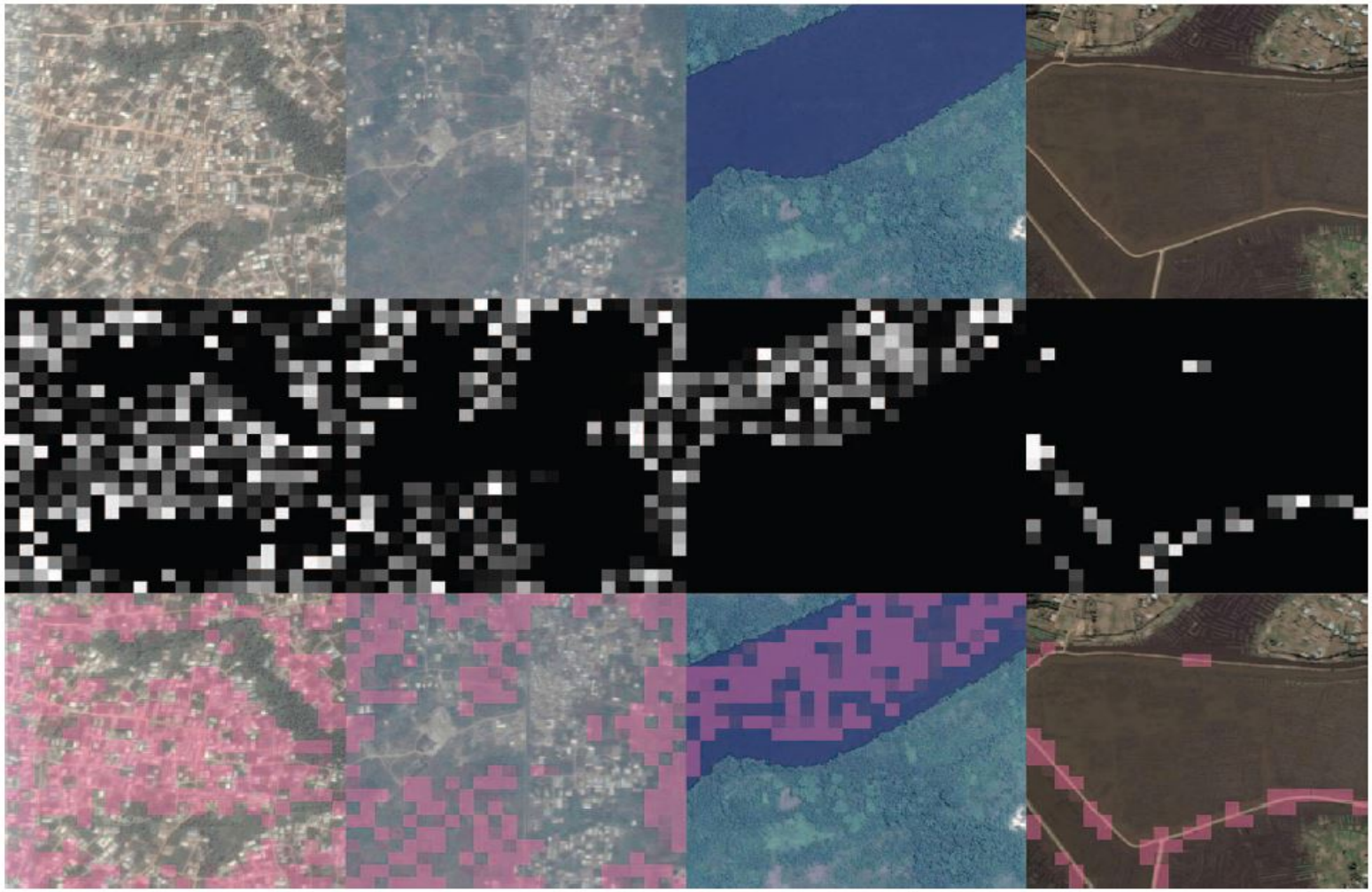

Figure: Four different convolutional filters in the convolutional neural network model used for extracting features. Source: Combining satellite imagery and machine learning to predict poverty(2016)

Data Sources

- World Bank;
- ZDHS
- Central Statistical Office, Zambia (Living Conditions Monitoring Survey Dataset)
- National Oceanic and Atmospheric Administration
- Google (Almost 17,000 images)

Expected Results

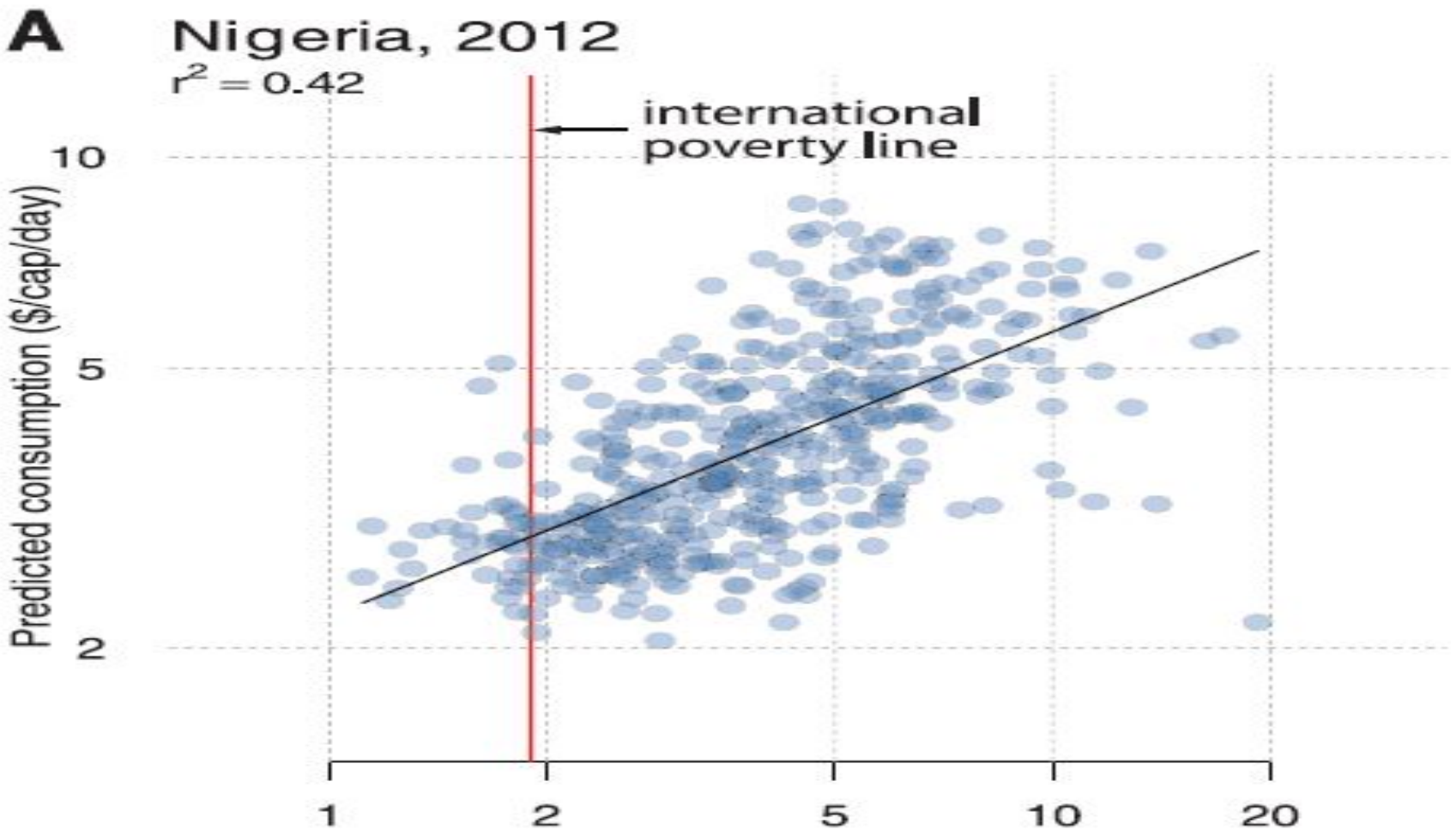

Figure: Expected Results.

Predictions and reported R-squared values from fivefold cross-validation

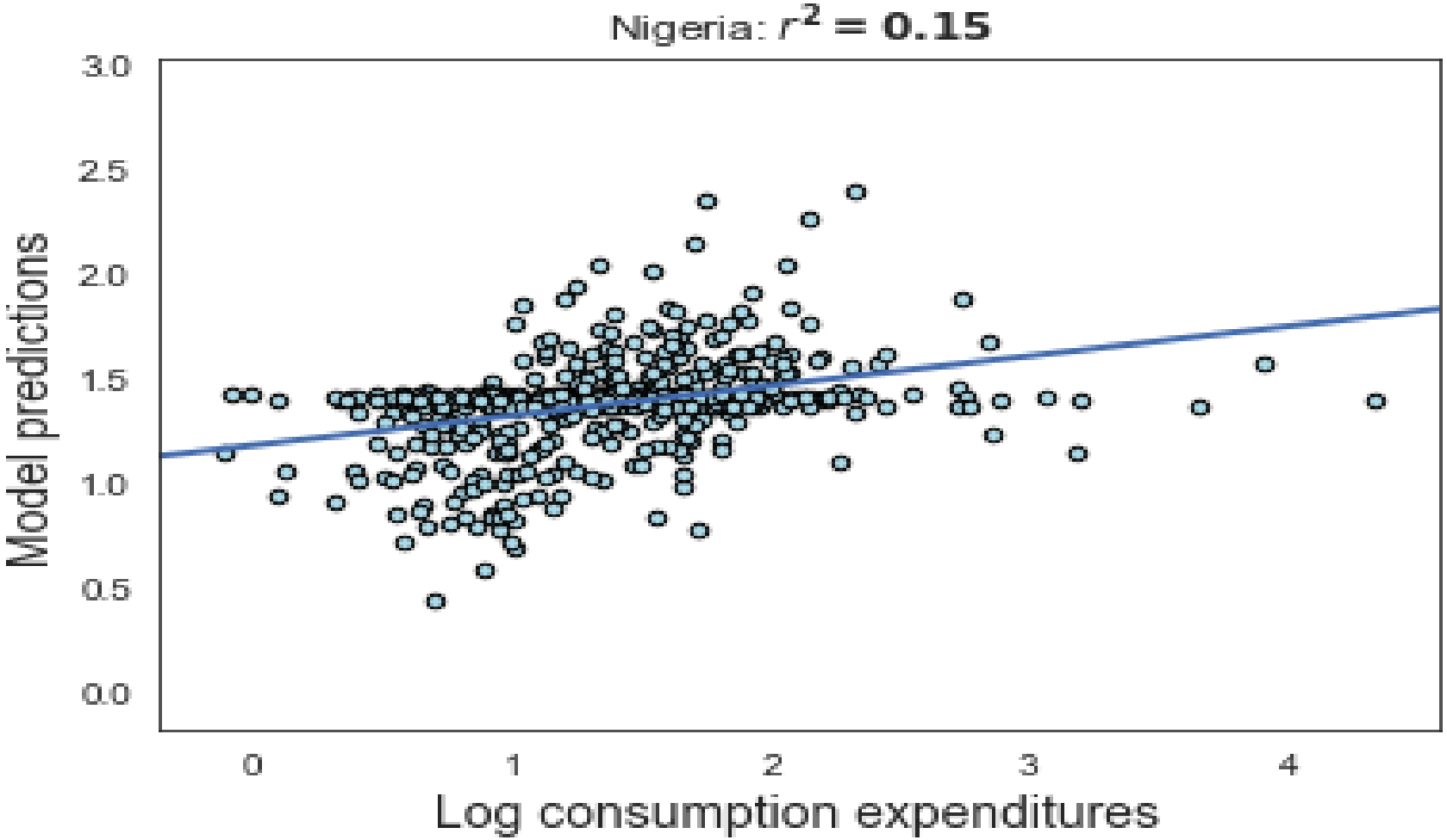

Figure: Expected Results.

Why are our Results Poor?

- We used Free Satelite Images and as such we could only obtain the 2019 Images ;
- Tunning was done on the 2016 Satelite Images

Next Steps

- Tune my own model ;
- Use newer data; the 2018 Dataset which will be available starting third week of January)
- Try to train tunning using WikiSatNet

REFERENCE

Thank You for your Attention