

# Pakistan - Poverty from the Sky

September 30, 2019

## 1 Introduction

We combine daytime and nighttime satellite imagery to predict poverty levels in Pakistan.

## 2 Data

Socioeconomic data comes from BISP, where we rely on 2011 and 2013 survey data which contains geocoordinates of household survey data. Our main variable of interest in the survey data is reported annual household income.

We test two sources of nighttime lights data: DMSP-OLS, which is available from 1992 to 2013 annually at a 1k resolution; and VIIRS, which is available monthly from 2012 to the present monthly at a 750 meter resolution. For daytime imagery, we rely on landsat which is available up to the present at a 30 meter resolution. Landsat data contains seven bands, representing different reflectance along the electromagnetic spectrum (e.g., reflectance of visible light – red, green and blue, and other bands such as infrared). The individual bands themselves do not provide an intuitive meaning; however, combinations of bands provide an intuitive meaning. For example, combining red and infrared bands provides a measure of the amount of vegetation.

As a simple approach to joining the household and satellite data, we take the average value of satellite data within a 2km buffer of each household.

## 3 Methods and Initial Results

We employ a series of machine learning models that use nighttime lights, individual daytime bands and combinations of daytime bands as input to predict household income. We implement a number of commonly employed methods, including support vector machines, gradient boosting and random forest. We employ a grid search to test a variety of parameters for each model type. We train the data on 70% of the sample and test on the remaining 30%.

The best models are decision trees which explain 70% of the variation in household income. We test models using daytime and nighttime features and just daytime features. Both sets perform similarly well; however, in models using all features nighttime lights dominates as the most important feature.