SDG 1: Ending Poverty in all its Forms

CSE 545 - Big Data Analytics

Team: SKAM 2022

Sudipto Bal (113219784), Keerthana Devadas (113260469), Ayush Dayani (114241407), Manoj Shet (113277283)

Introduction:

Poverty is a state of not being able to meet basic needs of living due to lack of money. However, poverty is not restricted to the lack of money. The World Bank Organization describes poverty in this way:

> *"Poverty is hunger. Poverty is lack of shelter. Poverty is being sick and not being able to see a doctor. Poverty is not having access to school and not knowing how to read. Poverty is not having a job, is fear for the future, living one day at a time."*

Going by the SDG, poverty is living with less than $3.50/day and extreme poverty refers to those who live on less than $1.90/day.

Currently, close to 9% of the population lives in extreme poverty and about 30% of people live in poverty. It is important to take measures to alleviate poverty as this is a huge number and cannot be ignored. In this project, we are aiming to - 1. Compare factors that affect people living in poverty, 2. Use night light satellite data to predict the poverty situation in 2030.

Background:

A lot of research and detailed study has been done in this field in recent years. Analysis of various factors affecting poverty is the best way to tackle poverty-related issues and take measures to eradicate the same. This comprises the first pipeline of our work.

Another approach we came across was aimed at predicting poverty in the year 2030 using night light satellite imagery. The basic assumption made here is that the more the light in the night sky for the area, the lesser the poverty. We found this direction intriguing and picked it to try a hand at it.

Data:

The main objective of the project was to analyse the night-time light intensity of a few developing countries to predict poverty, we collected satellite images of night-time light from the VIIRS database of Google Earth Engine. We scraped data from the VIIRS database for a region of South Asia, Africa, South America, and the USA. The total data collected was more than 40 gigabytes. We collected satellite images from 2014 to 2021 – one for each month. Thus, for each chosen geolocation, we had 97 images.

Region	**USA**	**South America**	**South Asia**	**Middle East**	**Africa**	**Australia**	**Europe**
Data in GB	7	4	7	3.6	10	2.4	5

Table 1. Image data size for each region

We also needed tabular data for the analysis of poverty-related factors. We got most of the data from 'Data World Bank' and 'World Bank Organisation' for 1990 – 2019, covering factors like education, health, sanitation, GDP, etc. This data accounted for 12 megabytes of the total dataset.

Methods:

We took two courses of action for the project, one for feature correlation and the other for predicting the night-time light related to poverty.

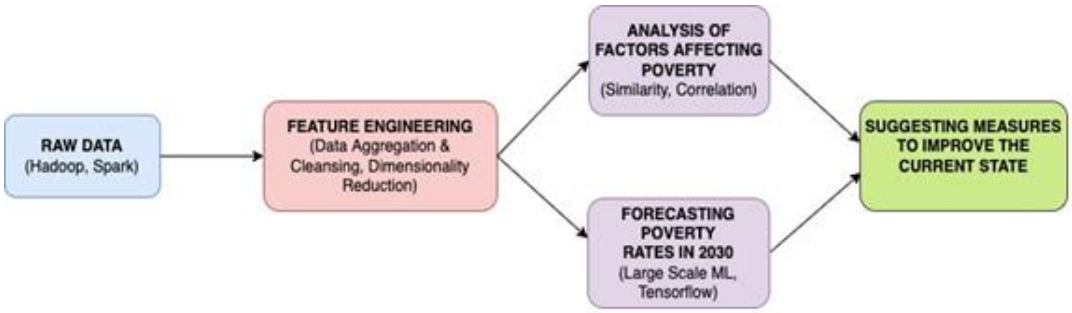

Fig 1. Process Flowchart

1. We processed the structured data (tables) with the use of spark. The data from different tables had to be aggregated in chronological order. After aggregation, the data was cleaned for incorrect, null, and redundant values. The cleaned data were used to derive Pearson correlations among features. Additionally, we tried to predict Poverty (< $3.50 a day) and Extreme Poverty (< $1.90 a day) from the current trends using ARIMA model up to the year 2030.

2. For the night-time light prediction, the images were of 1337 x 2228 dimensions. All 97 images for the geolocations were processed and the intensity of each pixel on the image was written into a csv file. The size of the csv file generated for each region is close to 8 gigabytes.

   Different parts of the image were processed, and the intensities of the pixels were predicted for the year 2030 using ARIMA model on different workers (distributed). The 'p', 'd', and 'q', parameters for the model were tuned on the fly for each pixel and the optimal value was taken. The result from the various workers was aggregated to form the final image. It was achieved using PySpark for data aggregation and TensorFlow for ARIMA model.

Evaluation/Result:

From the trends in the share of people living in poverty in countries since 1990, we have attempted to predict the conditions in 2030. This acts as a mirror to what the situation might be if the same level of effort is put into eradicating poverty and is maintained until 2030.

Given below in Fig 2 are examples of the predictions for a developing country (India) and an under-developed country (Uganda). We can observe that the current situation in India looks promising and it is on track to alleviate poverty, while the situation in Uganda is not too good.

The correlation analysis heatmap is represented in Fig 3. From this, we found very high correlations between factors as one would expect. We can see that Child Mortality, Unsafe Sanitation Deaths,

Malnutrition is very highly correlated with poverty, while Life Expectancy and Average Schooling Years are highly negatively correlated with poverty.

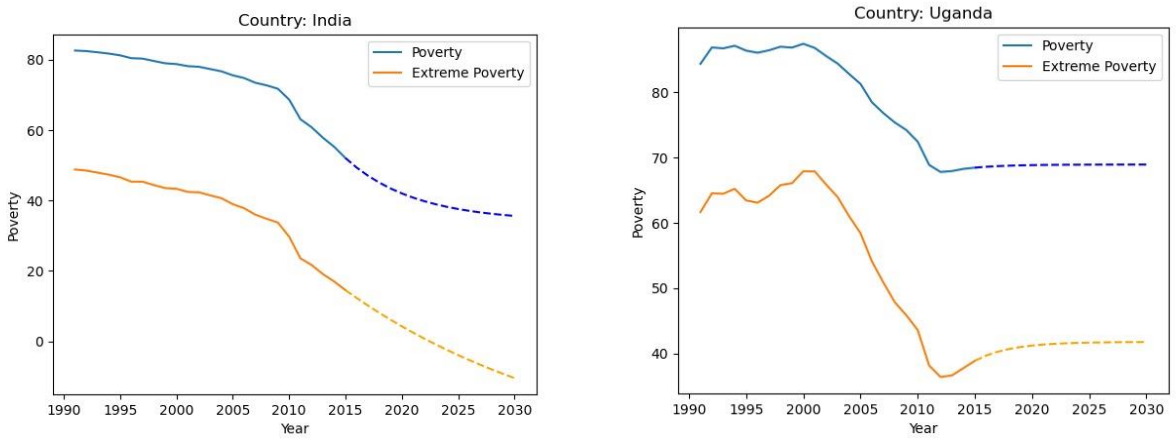

Fig 2 – Poverty predictions for India and Uganda

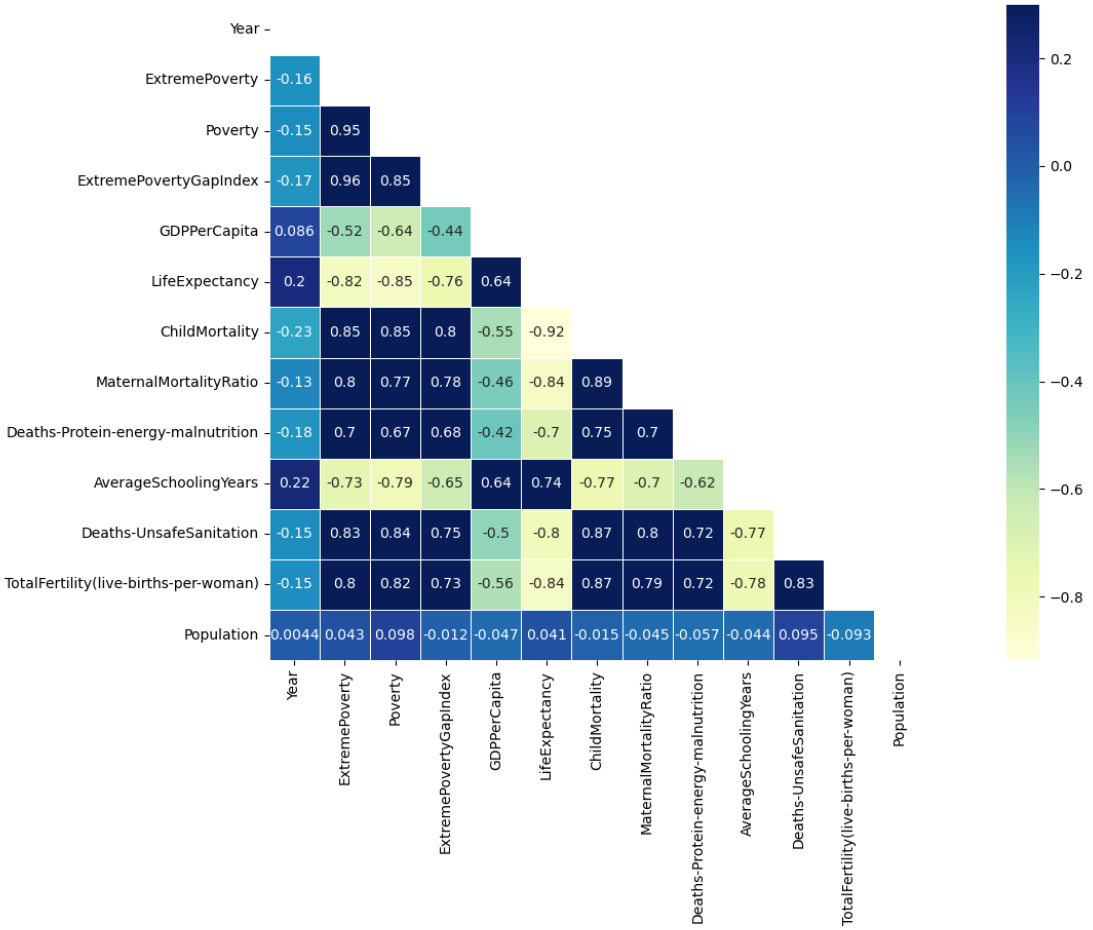

Fig 3. Feature correlations

Country	Poverty	Extreme Poverty
India	9.676278036	-11.11438878
Pakistan	20.20055108	4.368806811
Nepal	18.8660096	-0.502481901
Bangladesh	34.02532651	6.120810567
Bhutan	7.115132654	0.681173016
Thailand	-0.186116828	0.274350131
Myanmar	3.128511585	-0.435953552
Indonesia	-1.035402043	2.87823
Uganda	64.99622294	34.017347
Egypt	32.03615146	4.543123546
Namibia	36.09043243	15.69554875
Gambia	32.53446	7.977756
Zimbabwe	65.81351315	40.5525716
Niger	73.01192	39.29049
Nigeria	70.80816664	38.41422675
Angola	73.0806784	51.82883349

Table 1 – Predictions for 2030

We also predicted the nightlight images for a few regions covering South and South-East Asia. Due to time constraints, we could not complete execution for all the regions we had planned to, but the code can be used to do the same.

Here is an example of Southern India, first picture is from December 2013 and the prediction is for mid 2030.

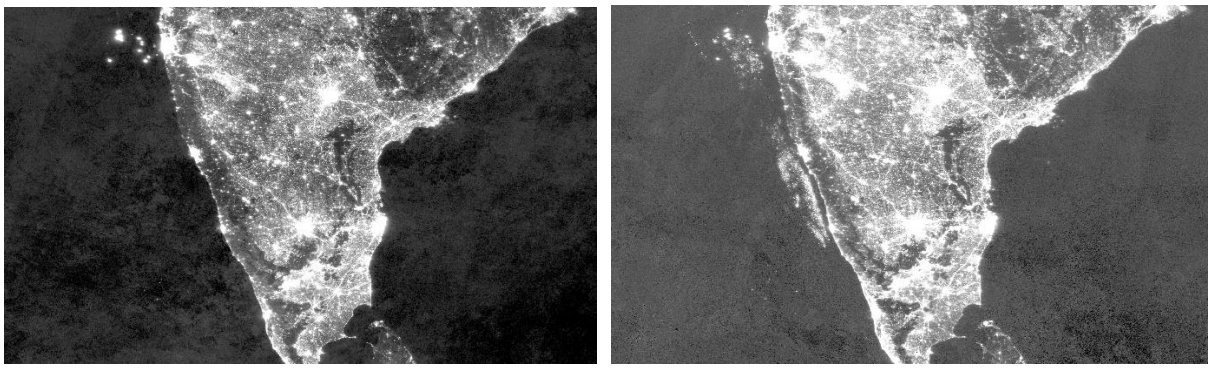

Fig 5 – Southern India in December 2013 (left) and 2030 (right)

Fig 6 – Northern Indian Subcontinent in December 2013 (left) and 2030 (right)

We can see that many areas on the land have turned brighter, especially those around already bright areas. A few issues we faced in this task are downloading data, snow cover and seasonal cloud cover over the area. Therefore, we can observe changes in pixel intensities over the sea.

Conclusion:

From our results and analysis, we can estimate if a given country is on track in its efforts to eradicate poverty and what factors one must focus on based on the correlations.

Some regions are dark due to the land being inhabitable, but there are also regions where there is no nightlight despite people residing. Normally, if people reside in an area and the region is dark, it is considered a sign of poverty. These areas can be identified from the predicted nightlight and must be more focused on.

Future scope: Better methods can be applied to the satellite imagery predictions like CNNs to get more reliable results.

References:

- https://data.worldbank.org/topic/poverty
- https://developers.google.com/earth-engine/datasets/catalog/NOAA_VIIRS_DNB_MONTHLY_V1_VCMCFG?hl=en
- https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/
- https://data.world/datasets/poverty
- https://ourworldindata.org/extreme-poverty
- https://ourworldindata.org/grapher/child-mortality-around-the-world?country=OWID_CIS~Less+Developed+Regions%2C+excluding+China~CHN~CHL~TCD
- https://ourworldindata.org/grapher/death-rate-from-unsafe-sanitation
- https://ourworldindata.org/health-meta
- https://ourworldindata.org/economic-growth
- https://ourworldindata.org/grapher/malnutrition-death-rates
- https://towardsdatascience.com/how-to-understand-global-poverty-from-outer-space-442e2a5c3666
- http://cs231n.stanford.edu/reports/2016/pdfs/423_Report.pdf
- https://worlddata.io/
- https://pubs.aeaweb.org/doi/pdfplus/10.1257/089533003769204335
- https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html
- https://geopandas.org/en/stable/docs/user_guide/mapping.html
- https://worldbank.github.io/OpenNightLights/wb-light-every-night-readme.html