Predicting hunger and food insecurity in sub-national administrative units across sub-Saharan Africa using weather, agricultural production and market price data

Team Members: Annie Phan (aqp), Matthew Mauer (mrmauer), Nixon Joshy (nixon)

The Food and Agricultural Organization (FAO) of the United Nations estimated that 820 million people suffered from hunger in 2018[1]. That is one of every nine people in the world did not have enough food to meet their minimum calorie requirements making them prone to stunting, cognitive impairment and susceptible to a variety of diseases. Failure to prevent hunger especially in childhood also has a cascading negative effect on educational attainment, health and overall quality of life[2]. The staggering scale of the problem had attracted attention across countries in the latter half of the 20th century leading to concerted global efforts to reduce global hunger. The very first Millennium Development Goal targeted halving the proportion of people suffering from hunger[3].

For decades the number of people living with hunger had been on the decline aided by strong economic growth across the developing world. However, this progress has now been halted with the number of food insecure people increasing annually since 2015. With the onset of COVID-19, the World Food Programme (WFP) predicts the doubling of the number of people suffering from acute hunger[4]. Thus, food security assumes renewed urgency and has taken center stage amongst all the Sustainable Development Goals[5].

To tackle global hunger, FAO, WFP, United States Agency for International Development (USAID), and other aid organizations make strategic decisions on where, when, and how to

---

[1] "Food | United Nations." https://www.un.org/en/sections/issues-depth/food/index.html. Accessed 4 May. 2020.
[2] "The Global Hidden Hunger Indices and Maps: An ... - NCBI." 12 Jun. 2013, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3680387/. Accessed 5 May. 2020.
[3] "MDG 1 - Eradicate extreme poverty and hunger - MDG Monitor." 15 May. 2017, https://www.mdgmonitor.org/mdg-1-eradicate-poverty-hunger/. Accessed 4 May. 2020.
[4] "Risk of hunger pandemic as COVID-19 set to almost double ...." https://insight.wfp.org/covid-19-will-almost-double-people-in-acute-hunger-by-end-of-2020-59df0c4a8072. Accessed 4 May. 2020.
[5] "Goal 2 - Sustainable Development - the United Nations." https://sustainabledevelopment.un.org/sdg2. Accessed 4 May. 2020.

deliver their limited supplies of food aid[6]. These decisions to allocate food aid are of limitless consequence to those directly suffering from hunger. There are a plethora of forces that may contribute to regional hunger or famine (drought, corruption, and war to name a few); crises may arise swiftly and demand policymakers to respond under strenuous time constraints. Any foresight into where and when the future famines will occur would enable aid organizations to lay the necessary administrative groundwork, establish partnerships, provide infrastructure, and marshall resources to effectively distribute aid to those with the greatest need while preventing it from falling into the hands of bad actors that may either steal or "tax" the aid.

The project aims to create a tool that can be used by such aid organizations to more efficiently plan and target their interventions by focusing on the regions in most need of aid. Predicting future episodes of food insecurity can help these organizations allocate food and plan for their distribution in advance of the event thereby reducing or eliminating its effects. The model's efficacy can be validated by testing it on data from 2019 using training data from the previous years to see if the model would have been satisfactory in its predictions.

Presently, the project plans to focus on countries situated in the Horn of Africa as a preliminary probe revealed availability of datasets.

Literature Review

There exist multiple studies linking the effects of hunger on human cognitive ability[7]. The challenge of global hunger is cited to be a well studied poverty trap[8] with those who suffer from it unable to make sufficient investments in themselves or their offspring to improve their human capital. Duflo et al. study the effects of undernutrition both in underdeveloped and developing countries to show how hunger results in a vicious poverty cycle[9].

The use of Machine Learning to predict incidences of food insecurity and hunger is an emerging use case that has attracted significant attention in recent years. Lentz et al. summarizes how the use of Machine Learning can improve global response to food insecurity crises using data on

---

[6] "Food systems | World Food Programme - WFP." https://www.wfp.org/food-systems. Accessed 4 May. 2020.
[7] "The Global Hidden Hunger Indices and Maps: An ... - NCBI." 12 Jun. 2013, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3680387/. Accessed 5 May. 2020.
[8] "Do Poverty Traps Exist? Assessing the Evidence - American ...." https://www.aeaweb.org/articles?id=10.1257/jep.28.3.127. Accessed 5 May. 2020.
[9] "Poor Economics - 3psmars." http://www.3psmars.org/wp-content/uploads/gravity_forms/1-56ca0c49093539c6a8bf145b1ef47cfa/2014/10/Poor-Economics.pdf. Accessed 5 May. 2020.

weather, agricultural prices and demographics[10]. The model developed by the researchers correctly predicted these crises with an accuracy of 83 to 99% depending on the measure.The World Food Programme has already begun exploring using Machine Learning for program monitoring and implementation[11]. There are other ongoing efforts to build similar models such as that by ICCO Cooperation[12]. These past efforts and their experiences shall be a guiding factor while building our model.

Data

Output Variable(s): The project aims to predict hunger using indicators to signify broader food insecurity. One prominent source for this indicator would be the food insecurity classification data by Famine Early Warning System Network (FEWSN)[13]. The data provides geographical information of the spread of existing food insecurity across Africa by classifying the extent of the crisis on a scale of 1 to 5. This rating is based on data, surveys and analysis done by humans and is provided on a quarterly basis. This is a potential output variable to predict. The Sub-national Food Consumption Score by WFP could be an alternative output variable.

Input Variables(s): The project foresees the utilization of weather, food price and conflict data to predict the output variable. The datasets to be used shall be merged on the geographical units of the output variable and trained for prediction. The datasets that shall be potentially used for training purposes are described as follows:

a. The World Food Programme publishes regular Food Price data for multiple geographic markets. For example, in Ethiopia it has published data[14] from 1992 on a monthly basis (sometimes biweekly) across 150 major markets for major staples.
b. There are multiple sources for daily weather data for weather stations across different countries. Data for African countries can be found using TAHMO[15]. However, the best and most prominent aggregator of weather data is the Global Historical Climatology

[10] "A data-driven approach improves food insecurity crisis ...." https://www.sciencedirect.com/science/article/abs/pii/S0305750X19301603. Accessed 5 May. 2020.
[11] "Unlocking Artificial Intelligence to beat Hunger | WFP Innovation." 15 May. 2017, https://innovation.wfp.org/blog/unlocking-artificial-intelligence-beat-hunger. Accessed 5 May. 2020.
[12] "Can Machine Learning Help Us Better Predict Hunger? - Icco." 22 Feb. 2018, https://www.icco-cooperation.org/en/blogs/can-machine-learning-help-us-better-predict-hunger/. Accessed 5 May. 2020.
[13] "FEWS NET Data Center | Famine Early Warning Systems ...." https://fews.net/fews-data/333. Accessed 4 May. 2020.
[14] "Ethiopia - Food Prices - Humanitarian Data Exchange." https://data.humdata.org/dataset/wfp-food-prices-for-ethiopia. Accessed 4 May. 2020.
[15] "African Climate Data | TAHMO." https://tahmo.org/__african-climate-data/. Accessed 4 May. 2020.

Network by the National Centers for Environmental Information[16]. The datasets found here include daily values for precipitation, snowfall, temperature, humidity and various other climatic indicators that could prove useful for our model.

c. Research shows that a strong indicator for predicting food insecurity is outbreaks of violence and conflicts[17]. ACLED publishes geolocated conflict and civil unrest event data with information on actors, dates, fatalities and types of events[18].

While the above three are the primary input datasets to be used, the following datasets may be integrated based on the model evaluation results.

d. Livelihood zones (similar to land use) published by Famine Early Warning System Network (FEWSN).
e. Sub-national government data on child stunting, malnutrition and prevalence of other allied conditions.
f. Sub-national government data on economic indicators such as income, wealth and savings.
g. Land use and vegetation cover data.
h. Demographic data published by sub-national administrative units.

Data Integration and Exploration

The first step of the project would be to merge the different datasets already available based on geographical units for which we have our output variable information - past incidences of hunger and food insecurity. The data integration and cleaning process is expected to take a significant amount of time in order to reach a final dataset where we have our input variables and the output variable merged showing the weather, market prices and conflict information during times of hunger.

Once the datasets are merged and cleaned a mix of visualizations and tests can be used to explore the data. Boxplot charts can be used to visualize the input variables such as rainfall, temperatures and market prices for major staples. These will also help in identifying outliers. Similarly choropleth maps for conflict and violence data can be a useful geographical comparison with the

---

[16] "Global Historical Climatology Network ...." https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/global-historical-climatology-network-ghcn. Accessed 4 May. 2020.
[17] "Food Insecurity and Conflict Dynamics: Causal Linkages and ...." 17 Jun. 2013, https://www.stabilityjournal.org/articles/10.5334/sta.bm/. Accessed 4 May. 2020.
[18] "ACLED." https://acleddata.com/. Accessed 4 May. 2020.

food insecurity maps published by FEWSN. Correlation maps as well as principal component analysis for the input variables could be a useful start for feature reduction and engineering.

Machine Learning

The project shall use input variables as described in the above section to predict the likelihood of hunger and food insecurity in a given sub-national administrative unit. The predicted variable will be transformed into a score ranging from 1 to 5 that signifies the likelihood and severity of a crisis over a given 3-month period. This is therefore a prediction problem which is likely to utilize regression to predict the scores. There may be some time series components involved to better predict these crises given the periodicity of such episodes — for example including the lagged outcome (food insecurity score) and a trending variable in our feature space .

The project currently aims to utilize Kernel Ridge Regression (KRR) with polynomial expansion of variables to predict the output variable. Alternative models such as random forests could also be utilized depending on the team's experience. Once market and conflict data are included, the feature space may be very large with some features highly correlated. Principal Component Analysis may provide additional insight in order to reduce dimensionality helping produce a simpler, sparser, more interpretable, and more accurate model.

Since the model works on quarterly output variables, speed of the model prediction will not be prioritized. The model may also lack external validity as it focuses on conditions found in the countries in Africa which may not translate to Latin America or South Asia. This is attributed to the unique cultural and socio-economic factors found in different parts of the world.

Evaluation

The dataset is expected to be imbalanced with most of the data points available for regions without hunger crises. To remedy this, multiple approaches including suitable penalties, oversampling for the regions with hunger crises or generating synthetic samples will be explored. Undersampling of the regions without hunger crises could also be done to achieve a more balanced training dataset. GridsearchCV will be utilized for tuning hyperparameters (e.g. normalization weights) and identifying the most effective kernel functions.

The model will be judged for its accuracy of correctly predicting hunger crises for the test data. Mean Absolute Error and Mean Squared Errors will be helpful towards this end. A confusion matrix for the transformed final scores ranging from 1 to 5 will help visualize the model's predictive abilities. Choropleths for the region/year of test year compared to the true food insecurity map for that year (since we shall keep aside the previous two years data as test data) would be another helpful visualization.

Ethics

The development of a machine learning based tool to solve any resource allocation problem is bound to raise questions on the fairness of redistribution. The model is meant to aid and not replace human judgement in making important aid decisions. The model also avoids making binary decisions for a more pragmatic scoring approach that can be one of the input factors playing into any human decision.